Molecular simulations and modeling of HIV-1 gp41 membrane spanning domain (MSD) in a model viral bilayer

A Thesis

Submitted to the Faculty

of

Drexel University

by

Michelle Katherine Baker

in partial fulfillment of the

requirements for the degree

of

Doctor of Philosophy in

May 2014

Dedications

My thesis is dedicated to my grandfather, Dayton Stout, who graduated from Drexel Institute of Art, Science and Industry in 1930 with a degree in engineering.

Acknowledgments

I would first like to thank my advisor Dr. Cameron Abrams for his encouragement and motivation. His dedication, enthusiasm, and quality of research have been a significant inspiration.

I would also like to thank my committee members Dr. Irwin Chaiken, Dr. Luis Cruz Cruz, Dr. Raj Mutharasan, and Dr. Masoud Soroush for their feedback and support.

I would also like to thank the Abrams' and Chaiken research groups including Dr. Rosemary Bastian, Anthony Bucci, Dr. Mark Contarino, Dr. Ali Emileh, Ryan Gordon, Dr. Changwoon Jang, Shyno Mathew, Dr. Francesca Moraca, Dr. Debashish Mukherji, Dr. Sergio Paz, Arun Sridhar, Dr. Spence Stober, Venkat Sundaram, Dr. Harish Vashisth, and especially Dr. Vamshi Gangupomu for their advice and camaraderie.

Finally, I would like to thank my family, Mom, Dad, Kim, Oz, and Mia, for their love and unconditional support.

Financial support from the National Institutes of Health (grant No. R01 AI-084117-01) is acknowledged. Anton computer time was provided under allocations No. PSCA10042P and PSCA13046P by the Pittsburgh Supercomputing Center (PSC), and the National Center for Multiscale Modeling of Biological Systems (MMBioS) through Grant P41GM103712-S1 from the National Institutes of Health. The Anton machine at NRBSC/PSC was generously made available by D.E. Shaw Research. This research was supported in part by the National Science Foundation through TeraGrid resources under grant No. TG-MCB070073N.

Table of Contents

List of Tables

List of Figures

Abstract

Molecular simulations and modeling of HIV-1 gp41 membrane spanning domain (MSD) in a model viral bilayer

Michelle Katherine Baker

Advisor: Cameron Abrams

HIV-1 envelope protein complexes known as "spikes" are trimers of gp120 and gp41 that mediate fusion and infection to target cell membranes. The membrane spanning domain (MSD) of gp41 contains several highly conserved residues important for fusion, however the structure and function of MSD are only partly understood. All-atom simulations can elucidate how the conserved residues affect MSD structure to provide atomistic insight into HIV-1 fusion.

Extensive molecular dynamics (MD) of monomeric HIV-1 gp41 MSD in model viral bilayers was used to investigate the conserved midspan arginine and the requirement of cholesterol for fusion. All wild-type peptides were α-helical, remained membrane-spanning, and solvated their midspan arginines with a water defect that was independent of cholesterol. However, the simulations indicate that cholesterol may allow the spike to localize the water defect and to control the tilt of the helices.

The dynamics of the model viral bilayer with $\sim$50% cholesterol was explored with 3 systems and simulated for up to 10 μs to explore the phase space of configurations, an order of magnitude greater than previous studies. This timescale allowed observation of diffusive motion and calculated diffusion coefficients agreed with experiments.

Oligomeric forms of the MSD were then created to examine the GXXXG motif, known for helical, transmembrane, dimer interactions, but conserved in the MSD of the trimeric gp41. The lowest-energy trimeric MSD with interacting GXXXG residues could not mediate trimerization in a bilayer on 100 ns timescales. However, the lowest energy MSD dimer remained associated on similar timescales, suggesting

a dimer form of the MSD during fusion.

Finally, the trimeric MSD was stabilized by addition of the trimeric crystal structure of the gp41 membrane proximal external region (MPER). Simulation of the MPER-MSD trimer for 11 μs showed relaxation towards a different, stable configuration in which the GXXXG motifs were not interacting but the cholesterol recognition motif (CRAC) sequestered water and cholesterol. This is the first simulation of a model of trimeric MPER-MSD in a cholesterol-containing bilayer and it may represent a point between the prefusion and the prefusion intermediate experimental trimeric structures.

1. Introduction

Since the discovery of acquired immunodeficiency syndrome (AIDS) in the early 1980s, much research has focused on the virus responsible, human immunodeficiency virus (HIV-1), but there is still no vaccine or cure. In 2013, the Joint United Nations Programme on HIV/AIDS (UNAIDS) reported that 35.3 million people are currently infected with HIV. In 2012, there were 1.6 million AIDS-related deaths, a 29% decrease since 2005, and there were 2.3 million new infections, a 33% decrease since 2001 [2]. The progress made by UNAIDS towards reducing the number of new infections, current infections, and AIDS-related deaths is a direct result of improved access and affordability of antiretroviral therapy (ART), which decreases viral loads and transmission. Unfortunately, ART and highly active antiretroviral therapy (HAART) are costly, have side effects, and are becoming less effective as the virus gains resistance. Therefore, new insights leading to novel therapeutics are strongly needed. Molecular details of the HIV-1 fusion and infection process are only partly understood. Structures of the fusion proteins gp120 and gp41, known collectively as "spike", are only partial and do not contain enough information to model the entire fusion process. In order to better understand the function of these proteins during fusion, more complete, unliganded structures in a membrane context are necessary. Structural investigations of spike are complicated because the fusion proteins are non-covalently associated and gp120 holds gp41 in a metastable structure in the native, unliganded spike. The gp41 protein mediates fusion of the viral and cellular membranes and is also a target for drug design. Partial structures exist for the more stable, soluble, postfusion gp41 six helix bundle (6HB) structure, but less is known about the prefusion, membrane-associated structures. HIV-1 infection requires cholesterol, however cholesterol in the viral membrane is not always accounted for in experiments.

Computational simulations can augment experimental investigations into the HIV-1 fusion process and the structure/function relationships of the spike proteins. This thesis will detail the use of all-atom molecular dynamics (MD) to characterize the membrane-associated sections of the HIV-1 envelope protein gp41.

1.1 HIV-1 lifecycle

HIV-1 fusion is mediated by envelope glycoproteins gp120 and gp41, which exist as a trimer of hetero-dimers or "spikes" on the viral membrane. HIV-1 particles have ∼10 spikes per particle, however it is not known if more than one spike is required for fusion [3]. By binding to CD4 on target cell membranes, gp120 initiates conformational rearrangements of spike, leading to co-receptor binding and to exposure of gp41 into its "prehairpin intermediate" state. This state involves bridging of the viral and cellular membranes; gp41's N-terminal fusion protein (FP) contacts the target cell membrane while gp41's membrane spanning domain (MSD) still anchors gp41 in the viral membrane. This prehairpin intermediate state rapidly folds into an extremely stable fusion state, where the heptad repeats of gp41 are arranged in a six helix bundle (6HB); it is thought that the formation of 6HB is the driving force for bringing together the viral and cell membranes and for fusion of the two bilayers. Infection occurs when the viral capsid containing the viral RNA enters the target cell through the fusion pore. The viral RNA is reverse transcribed into DNA and incorporated into a host chromosome. In order for production of new viral particles, the viral DNA is transcribed into mRNA, which is translated in the cytoplasm. (The mRNA for Env, the envelope proteins, is translated at the ER). Viral proteins are assembled at cholesterol-enriched sections of the cellular membrane. Env, after proteolytic cleavage of its precursor protein gp160, is incorporated into the cellular membrane also. The assembling virion starts to bulge from the cell membrane, aided by the polyprotein

Gag, and eventually buds off. As an enveloped virus, the HIV membrane is therefore "stolen" directly from cholesterol-enriched sections of the cellular membrane. After maturation, the mature viral particle is then ready to infect a new cell [4–7].

1.2 Importance of cholesterol in the HIV-1 lifecycle

The HIV-1 membrane is highly enriched in cholesterol, up to 45%, because it buds from microdomains in the cell membrane enriched in cholesterol [8]. A large amount of cholesterol in the viral membrane is required for infection. If cholesterol is extracted from the membrane using β-cyclodextrin (βCD), infectivity is abolished. Infectivity is quickly reestablished if the membrane is incubated with cholesterol-loaded βCD [9, 10]. Cholesterol also plays a role in other parts of the HIV-1 lifecycle, such as budding and co-receptor binding, but it is not known which properties of cholesterol are necessary for fusion [6, 11–15]. It is known that cholesterol increases the order and decreases the fluidity of bilayers [16]. Another possible mechanism of HIV-1 infection, compared to virus-cell membrane fusion, is the idea that the mature virion is osmotically stressed and that pressure on the spike can "pop" the virus, allowing the RNA-containing capsid to enter the cell membrane [17]. A stiff membrane, from a high amount of cholesterol, could relate to this idea of metastability of the spike and membrane.

1.3 Experimental structures of the spike

3D structure as determined by cryoelectron microscopy (cryo-EM) of the unliganded HIV or SIV spike shows a hollow cage-like ectodomain with either tripod or stalk membrane-adjacent structure for gp41 [18–20]. Partial structures of gp120 and gp41 have also been determined using X-ray crystallography and NMR [21–27]. The parts of spike that interact with the membrane are less known. A recent cryo-EM

structure of the uncleaved, unliganded HIV-1 spike includes part of the gp41 MSD as a left-handed α-helical coiled-coil with a crossing angle of 35° [28].

1.4 Segments of HIV-1 gp41

Dual functions for gp41 include anchoring the spike in the membrane and playing a role in fusion. The native structure of gp41 is unknown and it is postulated to be hidden and held in a metastable state by gp120. A "map" of the different sections of gp41 are shown below: starting from the N-terminal, FP is fusion peptide, FPPR is

N-	FP	-	FPPR	-	NHR	-	loop	-	CHR	-	MPER	-	MSD	-	CT	-C

fusion peptide proximal region, NHR is N-terminal heptad repeat, CHR is C-terminal heptad repeat, MPER is membrane proximal external region, MSD is membrane spanning domain, and CT is cytoplasmic tail. During fusion, the FP inserts into the target cell membrane and plays a role in perturbing the cell bilayer. It can adopt multiple membrane-associated structures, as it is conformationally flexible and seems to be highly influenced by its environment [29, 30]. The cholesterol content of the target cell membrane determines whether FP adopts an α-helical or β-sheeted structure [31]. The postfusion, soluble 6HB, composed of the heptad repeat sections NHR and CHR, has been characterized as an α-helical coiled-coil [21]. The 6HB has also been crystallized with extensions of FPPR and MPER [26]. Structures for MPER and MSD are discussed in the sections below. The CT contains a highly immunogenic region (HIR) and 3 α-helical lentiviral lytic peptides (LLP1, LLP2, and LLP3) that bind to the membrane on the endoplasmic side (inner leaflet). Interestingly, CT

is required for Env incorporation and fusion only in certain cell types (known as nonpermissive cell lines) [32].

1.4.1 HIV-1 gp41 MPER

Table 1.1: HIV-1 gp41 MPER peptide sequence

sequence	E	L	D	K	W	A	S	L	W	N	W	F	N	I	T	N	W	**L**	**W**	**Y**	**I**	**K**
residue index	660	.	662	.	664	.	666	.	668	.	670	.	672	.	674	.	676	.	678	.	680	.

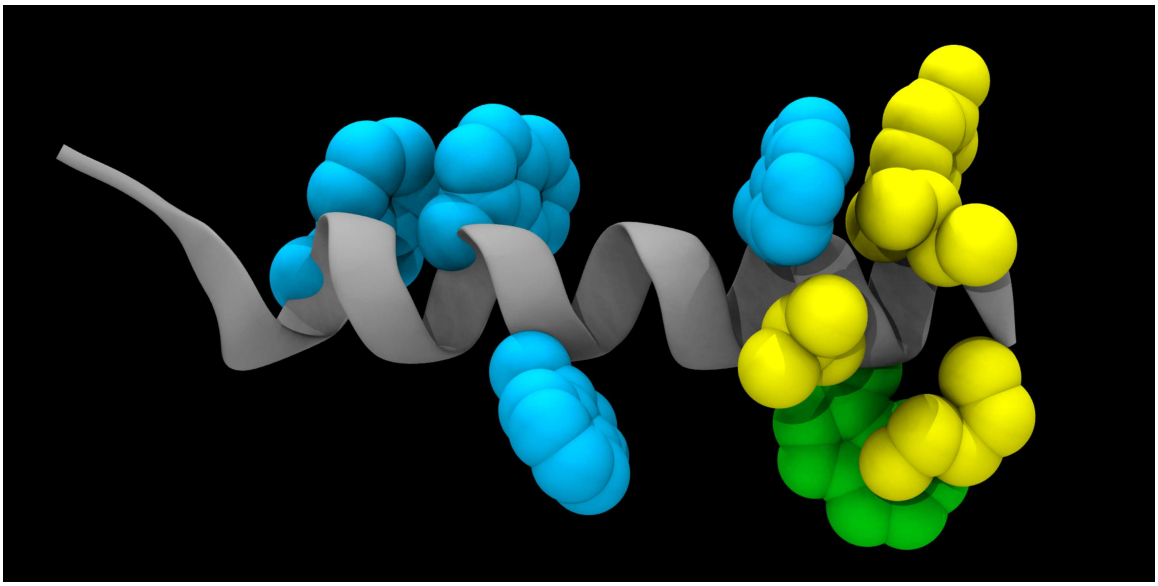

Figure 1.1: Representative MPER configuration with residues of interest highlighted: tryptophans in blue (W664, W668, W670, W676), CRAC domain in yellow (L677, W678, Y679, I680, K681), and CRAC tryptophan in green (W678).

MPER is important for partitioning of spike into cholesterol-rich domains before budding and for membrane perturbation during fusion. It has many highly conserved

residues, such as five tryptophan residues and a cholesterol-sequestering CRAC motif (LWYIK). The sequence of 22 residues that constitute MPER is shown in Table 1.1, with the CRAC residues in bold, and important residues are highlighted in Figure 1.1. Since MPER contains epitopes for the most broadly neutralizing antibodies against HIV-1, it has been a prime target for drug and vaccine design [33, 34]. However, the epitopes of MPER are only briefly exposed during fusion [25, 33, 35, 36]. Studies of full-length MPER, partial-length MPER, and CRAC domain peptides have revealed it has a high flexibility in a variety of environments (membrane prone, transmembrane, soluble) and has the ability to disrupt bilayers [23, 24, 37–43].

1.4.2 HIV-1 gp41 MSD

The membrane spanning domain (MSD) of gp41 is also dual-functional. Besides anchoring the spike, it has a specific sequence that is important for fusion. The functional importance of the specific sequence of the MSD is illustrated by experiments where replacement of the HIV-1 gp41 MSD with the MSD of the vesicular stomatitis virus G protein or influenza hemagglutinin glycoprotein abolished fusion to CD4+ HeLA or Jurkat T cells despite normal processing and expression [44, 45]. MSD has the most highly conserved residues important for HIV-1 infection, such as charged residues K681, R694, R705, R707, and the GXXXG motif, postulated to mediate transmembrane dimer interactions (see Table 1.2 and Figure 1.2). It is unknown if the GXXXG motif acts as a trimerization motif for the MSD, since gp41 is a trimer. Like other lentiviral MSDs, the HIV-1 MSD has a highly conserved midspan arginine, which places a positively charged residue in a hydrophobic bilayer [44, 46–48]. The single mutation of the conserved midspan arginine to leucine (R694L) was shown to have normal surface expression but diminished fusion activity [44, 49]. Mutations of the other conserved motifs showed similar consequences [46, 50–54]. It is not known

how the viral membrane accommodates a charged midspan arginine (R694) or why it is important for fusion. A "snorkeling model" has been proposed for the MSD, which would allow the midspan arginine to interact with negatively charged lipid headgroups [1, 47, 55]. Water defects have been observed in similar membrane systems with a charge, but the presence of a water defect has not been established in the HIV-1 gp41 MSD [56–64].

Table 1.2: HIV-1 gp41 MSD peptide sequence

sequence	K	L	F	I	M	I	V	G	G	L	V	G	L	R	I	V	F	A	V	L	S	I	V	N	R	V	R
residue index	681	.	.	684	.	.	687	.	.	690	.	.	.	694	.	.	.	698	.	.	701	.	.	704	.	.	707
local index	1	.	.	4	.	.	7	.	.	10	.	.	.	14	.	.	.	18	.	.	21	.	.	24	.	.	27

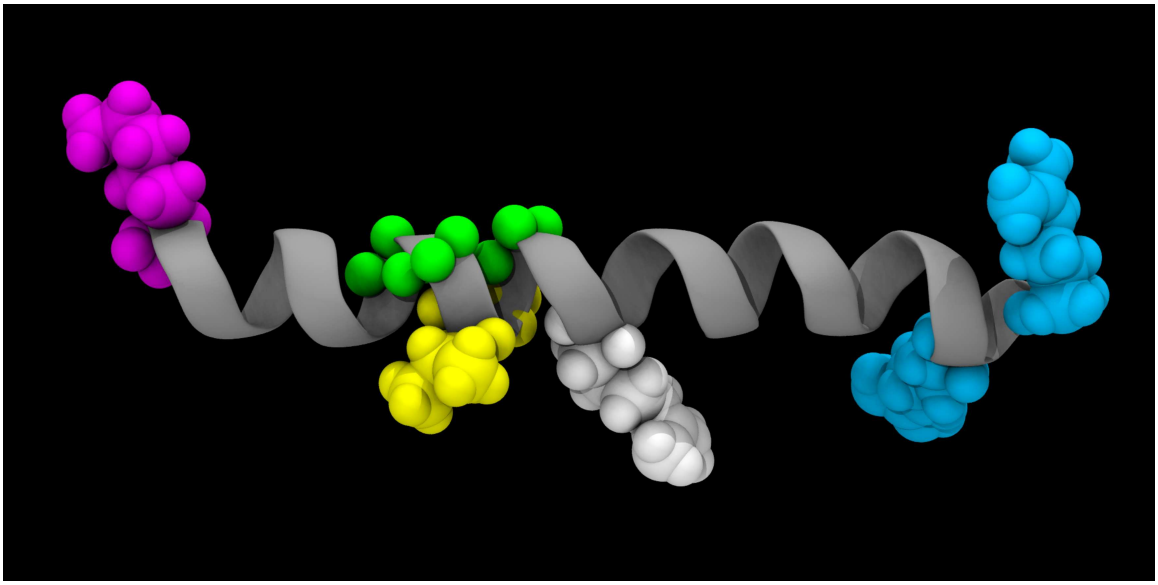

Figure 1.2: Representative MSD configuration with residues of interest highlighted: from left to right, lysine 681 in pink (L681), glycines in GXXXG motif in green (G688, G689, G692, first X is glycine), XX residues in GXXXG motif in yellow (L690, V691), midspan arginine in white (R694), and C-terminal arginines in blue (R705, R707).

Bioinformatic studies and circular dichroism (CD) analysis predict the MSD to be α-helical and mutagenesis experiments support that the MSD is membrane-spanning [54, 65–67]. However, there are no crystal or NMR structures of gp41 MSD yet, so it is difficult to understand the functional importance of the conserved residues. In general, membrane proteins such as gp41, are extremely difficult to experimentally determine the structure (by crystallography or NMR) and are underrepresented in the Protein Data Bank [68]. However, investigations into the structure of HIV-1 gp41 MSD in a model viral bilayer are appropriate for molecular simulations. This thesis details a segment-based approach to gp41 MSD structure determination using all-atom molecular dynamics. The structural narrative of gp41, especially the membrane spanning domain, is a key component to HIV-1 fusion and infection.

1.5 Aims of thesis

1.5.1 Aim I: HIV-1 gp41 MSD midspan arginine solvation

In order to construct a computational model of the HIV-1 spike in a cholesterol-containing membrane, segments of gp120 and gp41 can be individually studied and then combined together. In a first step towards building a native trimeric gp41, the monomeric MSD peptide in a model viral membrane was studied to elucidate the solvation of the midspan arginine and the potential influence of cholesterol on solvation.

1.5.2 Aim II: Dynamics of HIV-1 model viral membranes with cholesterol

Cholesterol plays many important roles in the HIV-1 lifecycle but it is not known which properties of cholesterol are important for fusion. The MSD of gp41 directly interacts with the viral membrane and a comprehensive study of the MSD must examine interactions with cholesterol and lipids. The dynamics of the model bilayer

simulated in Aim I were ascertained in Aim II. The composition of the model bilayer with cholesterol was evaluated for stability and relevance to HIV-1 infection.

1.5.3 Aim III: Modeling of HIV-1 gp41 MPER-MSD trimer in a cholesterol containing bilayer

The information obtained in Aims I and II was applied to construct a model of trimeric MSD in a cholesterol-containing bilayer. In Aim III, trimeric MSD models were enumerated in vacuum and the lowest energy structure was equilibrated in membrane. Trimeric MPER was augmented to the model and simulated on the μs timescale in order to see relaxation of the trimeric MPER-MSD model to a different configuration.

1.5.4 Aim IV: Role of the dimerization motif, GXXXG, in the MSD of trimeric HIV-1 gp41

Although it has been postulated that the conserved transmembrane dimerization GXXXG motif in MSD is responsible for MSD or gp41 trimerization, the motif disassociated in the 150 ns MSD trimer simulation and the 11 μs MPER-MSD trimer simulation in Aim III. Aim IV simulated the MSD dimer on the 100 ns timescale to identify possible reasons for trimeric disengagement in Aim III and to assess the validity of GXXXG as a trimerization motif for MSD.

2. Computational Methods

2.1 Molecular Dynamics

Molecular dynamics (MD) is a simulation technique based on classical mechanics where, given initial atomic positions and their potential, velocities and positions at the next point in time can be calculated. Repeating the calculation multiple times generates a trajectory, on the scale of ps to μs. The potential, in the form of a "force field", represents several interactions, including those of bonds, angles, dihedrals, van der Waals, and electrostatics. The integrator chosen for the calculations must provide the correct ensemble distribution, as dictated by statistical mechanics. For statistical mechanics theory and the principles of molecular dynamics, the reader is referred to references [69–72].

2.2 Protocols

All-atom MD simulations were performed with NAMD 2.8 and 2.9 [73] along with VMD 1.9 [74], the CHARMM force field [75, 76] with recent lipid-based corrections [77–79], and explicit TIP3P water. Unless noted, all systems were run in the NPT ensemble with a temperature of 310 K (body temperature) and a pressure of 1 atm. The NPT ensemble (constant number of atoms, pressure, and temperature) was simulated with Langevin dynamics with a coupling constant of 5 ps^{-1} for temperature control and with the Langevin Nosé-Hoover method for pressure control [80, 81]. The pressure was treated isotropically as fluctuations in x and y (corresponding to membrane lateral area) are uncoupled to fluctuations in the z direction in membrane simulations. Electrostatics were accounted for by the particle-mesh Ewald (PME) method [82] for systems with periodic boundary conditions with a grid spacing of

either 1 or 2 Å. Verlet integration was applied every 1 or 2 fs. Nonbonded forces were shifted off from either 8 to 9 Å or 10 to 12 Å.

2.3 Previous Group Work

Previously, the initial coordinates for gp41 MSD were created by starting with an NMR structure of a 27 amino acid α-helical subdomain of HIV-1 Vpu protein from the Protein Data Bank [83]. Using those initial positions for the backbone, the side chains of Vpu were removed and the side chains of HIV-1 gp41 MSD were added using VMD. The peptide was solvated, neutralized, ionized to a strength of 0.1 M (Na^+ and Cl^-), and equilibrated (100 ps NPT, 20 ns NVT) before being pulled into a 512-lipid membrane using constant velocity steered molecular dynamics (SMD), as rendered in Figure 2.1. Two different conformations were created by pulling either the N- or C-terminus, leading to an endoplasmic snorkeling R694 and an exoplasmic snorkeling R694, respectively. The initial membrane coordinates were created with the CHARMM-GUI membrane builder and contained 50% dipalmitoylphosphatidylcholine (DPPC) and 50% cholesterol by mol [84]. The membrane was solvated with explicit TIP3P waters, neutralized, ionized, and equilibrated for 6 ns NPT MD before the addition of the peptide. During SMD, the lipid headgroups were restrained in the z-direction and the steering velocity was 10 Å/ns. For the exoplasmic snorkeling model, the SMD atom was the center of mass of R707. For the endoplasmic snorkeling model, the SMD atom was the center of mass of K681 and the secondary structure was restrained. The peptide-membrane system was cut to a smaller size, neutralized, ionized to 0.1 M, and equilibrated for 5 ns NVT MD [85]. The resulting system contained 151 lipids and 183 cholesterol molecules (resulting in a membrane composition of 45.2% DPPC and 54.8% cholesterol) and was then used to initiate both long production MD and metadynamics simulations.

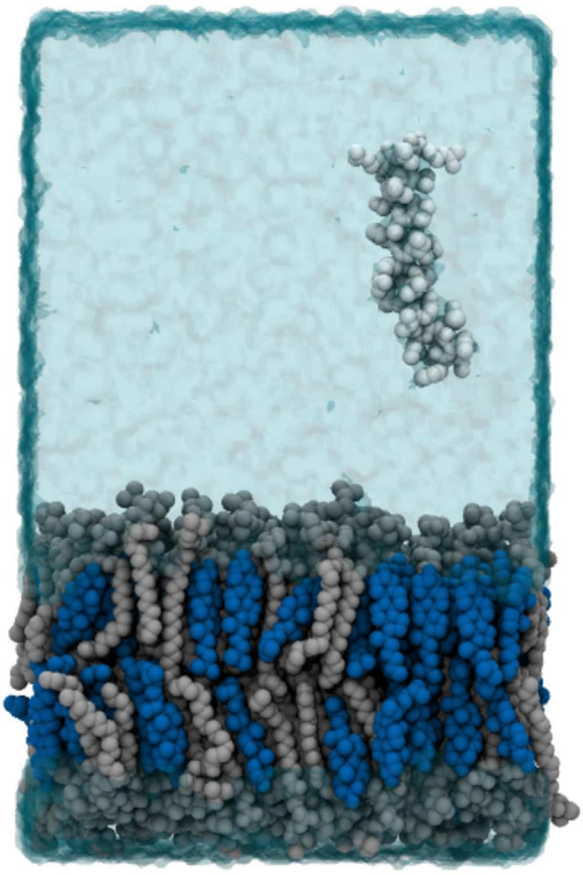

Figure 2.1: Initial configuration for SMD. Lipids are in silver vdW, cholesterol in blue vdW, protein in white vdW, and water in cyan isosurface. No hydrogen atoms are shown for simplification.

2.4 Choice of Lipids

The MSD peptide was embedded in a bilayer composed of ∼50% cholesterol and ∼50% DPPC, which represents a model of the HIV-1 viral membrane. Since new HIV-1 particles bud from sections of the cell membrane high in cholesterol, the HIV-1 bilayer is about 50% cholesterol [8]. Another major component of the viral bilayer is sphingomyelin (SM). However, the force field utilized in these simulations, CHARMM, did not have any parameters for SM in the open literature. (However, one month af-

ter my defense, CHARMM parameters for sphingomyelin were published [86].) A common computational and experimental substitute is DPPC. Both DPPC and SM are fully saturated lipids and undergo their main phase transition around the same temperature [87, 88]. Importantly, there exist CHARMM parameters for DPPC. Therefore, the binary model of ∼50% DPPC and ∼50% cholesterol is an acceptable, but simple, model of the HIV membrane. At the physiological temperature of 310 K, this binary model will be liquid ordered [89, 90], which is confirmed by lipid order parameters discussed in Chapter 3. The model used here of a cholesterol-free virus is 100% DPPC. There does not exist in literature, to the best of our knowledge, a measurement using MD of the DPPC phase transition temperature with the CHARMM36 force field, but other force fields calculate T_m to be between 300 and 320 K [91, 92]. This compares fairly well to experimental measurements of 314-315 K [93, 94]. Indeed, the pure DPPC bilayers simulated here are gel-like and, according to lipid order parameters, are slightly less ordered than the cholesterol-free bilayers (see Chapter 3).

2.5 Benchmark Parameters

Benchmark parameters for several representative systems are located in Table 2.1.

2.6 Trajectory Analysis

The trajectories were analyzed with codes/scripts in *c*, *tcl*, and MATLAB. (See Appendices A and B for examples in *tcl*.) Figures were either rendered in VMD [74] or created using gnuplot or GIMP.

Table 2.1: Benchmark parameters for eight representative simulations

System	Simulation Time	# of atoms	MD parameters	Version of MD Software	Supercomputer	# Cores	ns/day	cores x time per step (s)
WT1	300 ns	58,185	1 fs, PME 2.0, nb 8 to 9 Å	NAMD 2.8b1	Lonestar	192	42.7	0.4
WT2	300 ns	59,288	1 fs, PME 2.0, nb 8 to 9 Å	NAMD 2.8b1	Lonestar	192	41.0	0.4
WT1c	100 ns	58,185	2 fs, PME 1.0,nb 10 to 12 Å	NAMD 2.9	Stampede	96	15.5	1.1
WT2ΔChol	300 ns	89,017	2 fs, PME 1.0,nb 10 to 12 Å	NAMD 2.9	Stampede	96	10.4	1.6
R694L Metadynamics	320 ns	57,948	1 fs, PME 1.0, nb 8 to 9 Å	NAMD 2.8b1	Ranger	96	15.7	1.1
MSD+MPER trimer	133 ns	97,825	2 fs, PME 1.0,nb 10 to 12 Å	NAMD 2.9	Stampede	96	9.1	1.8
WT2	8 μs	59,381	2 fs, see Appendix C	Anton 2.6.4	Anton	512	3588	2.5e-11
MSD+MPER trimer	12 μs	97,825	2 fs, see Appendix C	Anton 2.12.4	Anton	512	3707	2.4e-11

3. Aim I: HIV-1 gp41 MSD midspan arginine solvation

3.1 Introduction to Aim I

The membrane spanning domain (MSD) of HIV-1 gp41 anchors the fusion complex known as "spike" in the viral membrane. It has a specific amino acid sequence that is highly conserved and is important for fusion/infection. For instance, the positively-charged midspan arginine (R694, residue 14 of 27) exists in 98% of sequences in the Los Alamos HIV-1 database [95] and mutations of the arginine negatively impact fusion [49]. Since the MSD is postulated to be helical and to span the bilayer once, this potentially places R694 in the middle of a hydrophobic environment. A "snorkeling" model has been proposed in which R694 can access lipid headgroups and water [1, 47, 55], perhaps creating a "water defect". Water defects have been observed in both simulations and NMR experiments on charges in a hydrophobic environment, such as guanidinium ions, arginine analogs, and transmembrane proteins with arginine(s) in a lipid bilayer [56–64]. In fact, the free energy of insertion of an arginine in a bilayer has been calculated to be a few kcal/mol with solvation [59, 96, 97]. It is reasonable to think that R694 of MSD can cause water ingress, although it has not been previously studied. Therefore, in Aim I we considered whether MSD had a water defect, considered how such a defect depended on the membrane composition, and speculated on the role of such a defect in the fusion process. Also, the HIV-1 viral membrane has a high cholesterol content; cholesterol seems to have multiple roles in the HIV-1 lifecycle and cholesterol depletion from the virion compromises fusion [8, 11, 13]. Starting with a model of the monomeric MSD in a cholesterol-containing bilayer, we characterized the equilibrium properties of MSD on a 300 ns timescale using all-atom molecular dynamics (MD). We utilized multiple replicas,

target mutations, protein-free membranes, cholesterol-containing and cholesterol-free membranes, and 3.4 μs total simulation time to systematically describe in detail the protein and membrane properties of the system. (See Figure 3.1 for some representative systems.) We aimed to answer the following questions:

1. What mechanism allows for the midspan arginine to compensate for its charge and how does this affect the MSD configuration?
2. How does the presence or absence of cholesterol affect the MSD configuration?
3. Does cholesterol influence the midspan arginine solvation?

3.2 Methods

3.2.1 System setup

The initial all-atom system consisted of the 27-residue MSD peptide in a bilayer (composed of ∼50% dipalmitoylphosphatidylcholine (DPPC) and ∼50% cholesterol) with explicit water and ions; this system was known as WT1. Two more replicas were created by pulling the MSD peptide with constant velocity SMD into different locations in the initial model bilayer, resulting in different local lipid arrangements around the MSD (see Appendix Figure 3.14). These systems with different initial conditions were referred to as WT2 and WT3 and were equilibrated for 6 ns NPT MD and 1 ns NVT MD before being subjected to equilibrium sampling. In order to investigate the importance of the midspan arginine, R694 in the WT1 system was mutated to leucine by using the Mutator plugin in VMD. Leucine is a hydrophobic residue that should not be solvated. This mutant system was referred to as the R694L system and after mutation, was further equilibrated for 3 ns in the NPT ensemble and then run for 5 ns in the NVT ensemble. To remove any initial condition bias

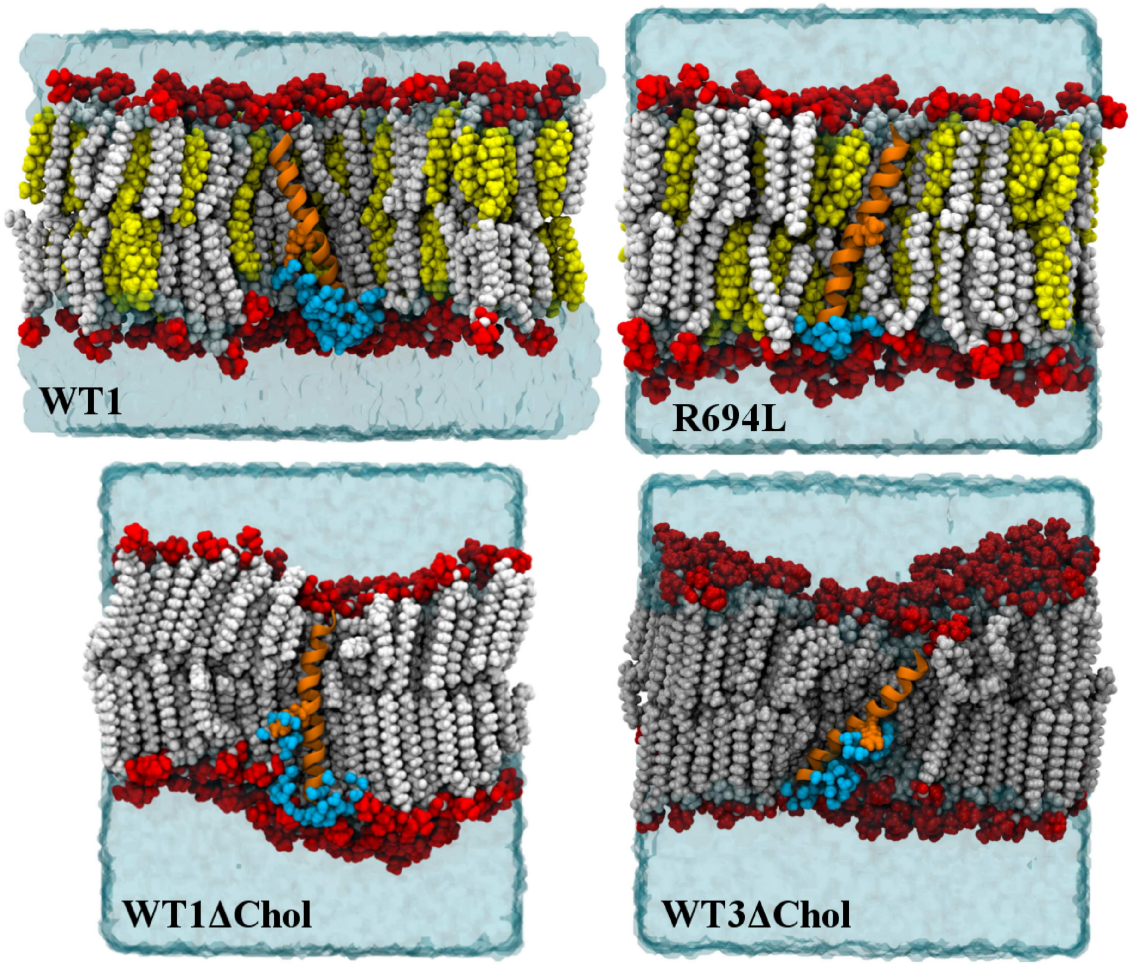

Figure 3.1: Representative system configurations from equilibrium MD at 300 ns rendered in VMD for the WT1, R694L, WT1ΔChol, and WT3ΔChol systems: lipid headgroups are in red, lipid tails in white, cholesterol in yellow, water in cyan, peptide in orange, and R694 or R694L in orange spheres. Water with oxygens within 4 Å of the protein and located on the inner leaflet of the membrane are shown in cyan spheres. For clarity, lipids, cholesterol, and water in the foreground of all four configurations are not shown.

associated with pulling the WT peptide into the membrane, the leucine was mutated back to arginine (R694LR) after the R694L system was simulated for 300 ns. In order to investigate the importance of cholesterol, the MSD peptide was also pulled into a pure DPPC bilayer using constant velocity SMD and equilibrated, resulting in the WT1ΔChol system. Two more replicas were also created (WT2ΔChol and WT3ΔChol); the WT1ΔChol system was branched off at pre-equilibration and 0

ns, the protein backbone was constrained, and the systems were simulated for 25 ns NPT MD to allow further equilibration of lipids and water around the proteins before production MD. Protein-free systems were referred to as DPPC and DPPC/Chol systems. For production runs, all systems were simulated for 300 ns NPT MD. The cholesterol-containing systems had ∼58,000 atoms, including 183 cholesterol and 151 DPPC molecules. These systems were roughly 80 x 80 x 80 $Å^3$. The pure DPPC systems had ∼89,000 atoms, including 334 lipid molecules. These systems were roughly 90 x 90 x 100 $Å^3$. A few of the systems (WT1, R694L, WT1ΔChol, and WT3ΔChol) are shown in Figure 3.1.

3.2.2 Comparison of parameters on accuracy and performance

The initial systems WT1 and WT1ΔChol were first simulated for 100 ns with different MD parameters to assess computational time and accuracy. The first set of parameters were based on those recommended by CHARMM-GUI Membrane Builder (grid spacing of 1 Å, 2 fs timestep, nonbonded shifted from 10 to 12 Å) [84]. The second set was more aggressive (grid spacing of 2 Å, 1 fs timestep, nonbonded shifted from 8 to 9 Å). For the cholesterol-containing systems, both the CHARMM-GUI-recommended and more aggressive parameters resulted in negligible differences in observables for the two systems (WT1c and WT1, respectively) as seen in Appendix Figure 3.15. Therefore, all cholesterol-containing systems were run with the second set of parameters for the slightly lower computational time. However, the cholesterol-free systems had to be run with the first set of parameters to maintain accuracy of certain observables. See Table 2.1 for benchmark parameters for WT1 and WT1c.

3.2.3 Metadynamics

The conformational space of the model of the WT MSD in a bilayer was previously sampled to assess the stable state for equilibrium MD [1]. In order to sample the conformational space of the mutant peptide, metadynamics was used to accelerate rare conformational events and construct the free energy profile along a particular CV. In metadynamics, the MD trajectory is biased by a history-dependent potential that is the sum of Gaussians deposited along the CV; the forces arising from these Gaussians allow high-probability regions along CV space to be explored first and discourage the system from revisiting the same CV values [98–101]. From Laio et al., the external potential is defined as

$$V_{\text{meta}}(\xi) = \sum_{t'}^{t'<t} W \prod_{i=1}^{N_{\text{cv}}} \exp\left\{ -\frac{[\xi_i - \xi_i(t')]^2}{2\delta_{\xi_i}^2} \right\}, \tag{3.1}$$

where ξ_i is the value of CV, $\xi_i(t')$ is the value of ξ_i at time t' from the atomic coordinates, W is the Gaussian height, δ_ξ is the Gaussian width, and t' values are the frequency in MD steps with which Gaussians are deposited. The R694L system was started from the NVT equilibrated, mutated system and run for 320 ns of constant-NVT metadynamics, after indication that the system had become diffusive, with a timestep of 2.0 fs and all hydrogen bonds constrained by SHAKE [102]. The CV was defined as the RMSD of the backbone atoms (C, C_α, N, O) of the peptide with respect to a perfect α-helix, an average over a total of 109 atoms. The lower boundary, an RMSD of 0 Å, indicated a perfect helix, and the upper boundary, an RMSD of 10 Å, indicated a partially unfolded peptide. The parameters included a Gaussian width of 0.01 Å, a weight of 0.15 kcal/mol, and a deposition frequency of 1 ps. Standard MD was continued after 300 (CV of 4.8 Å) and 304 ns (CV of 8.7 Å) of metadynamics for 20 and 37 ns, respectively, in the NVT ensemble to confirm the stable state.

3.2.4 Calculations of observables

The mass density of water (within 4 Å of protein and global), lipid headgroups, lipid tails, cholesterol, protein, and midspan residue in amu/Å^3 along membrane normal from the last 100 ns was calculated by first rewrapping the trajectories in x and y by using the peptide center of mass and then in z by using the bilayer center of mass. Statistics for WT include systems WT1, WT2, and WT3 and statistics for WTΔChol include WT1ΔChol, WT2ΔChol, and WT3ΔChol. Local density of lipid headgroups, lipid tails, and cholesterol (within 6 Å of protein) in amu/Å^3 along membrane normal was also calculated from the full 300 ns trajectory. The tilt of the helix was determined by measuring the angle between membrane normal (z-axis) and a vector aligned along the helix. The number of unique water molecules within 4 Å of the protein was attributed to the closest residue (Appendix A). To create 2D membrane thickness maps, the bilayer was sectioned into 4 x 4 Å^2 patches and the difference between the centers of mass of the lipid headgroups in each leaflet was attributed to a patch. In the lower leaflet, the average minimum water distance from the bilayer interface was calculated for each 4 x 4 Å^2 patch to create the 2D water penetration maps. Lipid order parameters, S_{CD}, for each DPPC tail were calculated from

$$S_{CD} = \langle \frac{3\cos^2\theta - 1}{2} \rangle, \tag{3.2}$$

where θ is the angle between the carbon-hydrogen vector in a lipid acyl chain and bilayer normal [103]. Hydrogen bonds were calculated using a donor-acceptor cutoff of 3.5 Å and an acceptor-donor-H cutoff of 30° [61].

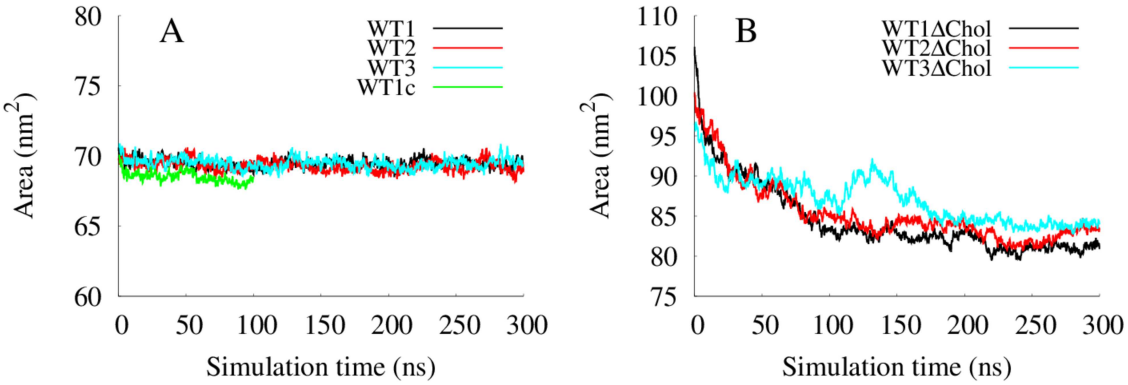

Figure 3.2: Unit cell area in nm^2 vs simulation time in ns for A) cholesterol-containing systems WT1, WT2, WT3, WT1c and B) cholesterol-free systems WT1ΔChol, WT2ΔChol, and WT3ΔChol.

3.3 Aim I results

3.3.1 Model membranes were stable during 300 ns MD

Both the cholesterol-containing and cholesterol-free membranes were stable on the 300 ns timescale and had properties that were consistent with previous experimental and computational measurements. Figure 3.2 shows the unit cell area in nm^2 during the trajectories of several systems. The cholesterol-containing systems in Panel A equilibrated quickly and had a stable unit cell area of around 70 nm^2. Also, the WT1c run with different simulation parameters as discussed in Methods had a similar area to the production WT systems. In Panel B, the cholesterol-free systems required over 100 ns to equilibrate to a stable area of around 83 nm^2. Apparently, the initial membrane configurations were not stable and needed over 100 ns to reach a more stable state. However, the average area per lipid for the DPPC system after 300 ns, 48.8 Å^2, compared fairly well with the gel-phase experimental value of 47.2 Å^2 [104]. The area per lipid for the systems in Panel A is not trivial to calculate since the systems contained two lipid components and a protein. However, other parameters

like thickness, discussed in Section 3.3.5, and lipid order parameters, discussed below, were consistent with previous experiments on model membranes with the same composition.

The lipid order parameters, S_{CD}, for both model membranes also supported that the models were stable and consistent with liquid ordered (for cholesterol-containing) or gel (cholesterol-free) bilayers [91, 105]. As seen in Appendix Figure 3.16 A and B, the cholesterol-containing systems had high order parameters, which indicate that they are very ordered.

3.3.2 Conformational sampling of HIV-1 gp41 MSD mutant with metadynamics

The single mutation of the conserved midspan arginine to leucine (R694L) was shown experimentally to have normal surface expression but diminished fusion activity [44, 49]. Differences in system equilibrium properties (such as mass density, peptide tilt, membrane thickness, and water organization) between the WT and fusion/infection-diminished R694L MSD systems can highlight peptide properties or molecular interactions that are favorable for viral membrane fusion and infection. The effect of this mutation on the peptide structure and on the membrane environment is not known as no experimental results on this question exist, to our knowledge. Although it is not likely that the single mutation significantly changed the secondary structure of the R694L MSD, it was still investigated for comparison to the previous study on the WT MSD [1]. Conformational sampling of proteins using MD is limited because accessible computation times do not permit rapid and repeated crossing of free energy barriers in conformational space; therefore enhanced sampling methods that invoke low-dimensional collective variables (CVs) and restricted free energies are often advantageous. So, in order to confirm that the R694L MSD was also α-

helical, metadynamics was conducted on the R694L MSD in a ∼50% DPPC and ∼50% cholesterol bilayer. This technique allowed us to reconstruct the potential of mean force (PMF) along a conformational collective variable (CV), the RMSD of the peptide backbone as compared to a perfect helix, and to then determine the stable state [98–101].

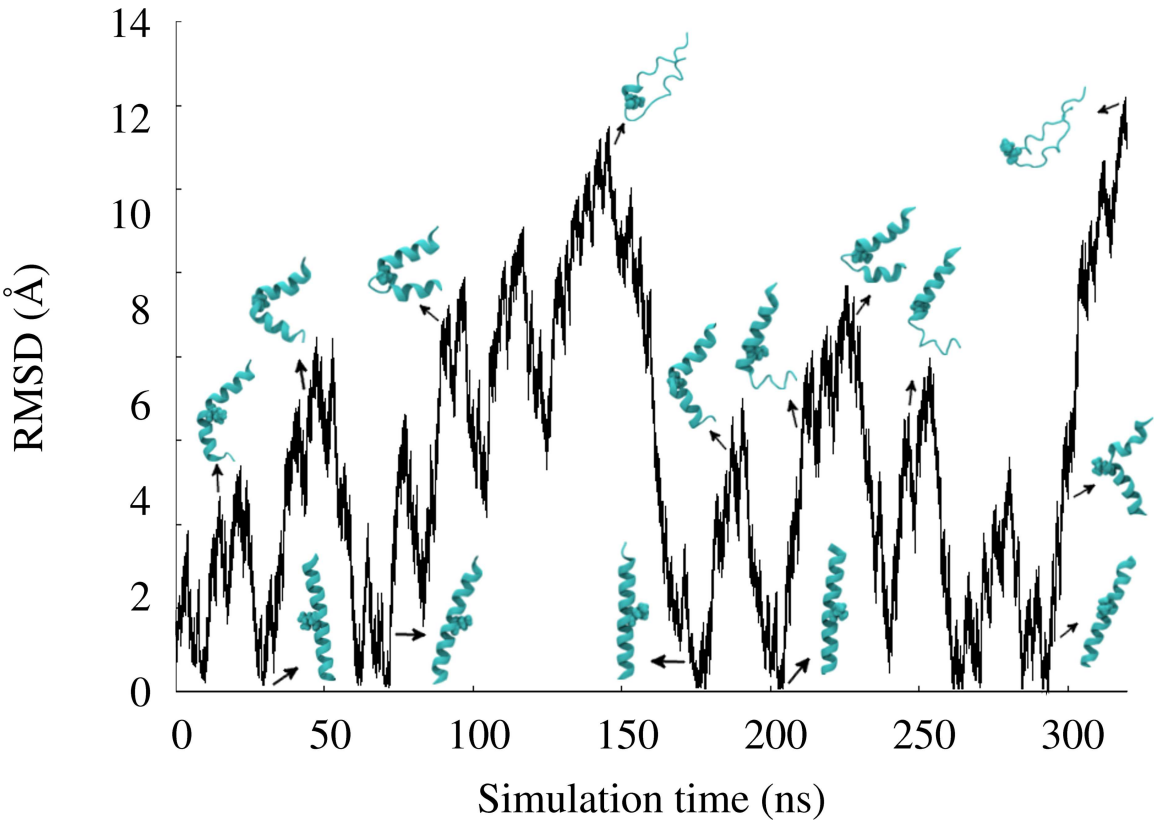

Figure 3.3: Collective variable (RMSD of backbone compared to perfect helix) in Å vs simulation time in ns for R694L peptide metadynamics. As insets are snapshots from along the R694L peptide metadynamics trajectory rendered in VMD: R694L peptide in cyan new cartoon and R694L residue in cyan vdW.

Metadynamics successfully explored the conformational space of the R694L MSD in 320 ns, at which point the CV became diffusive over the CV range explored (Figure 3.3) and the cumulative potential of mean forces (PMFs) of the last 20 ns con-

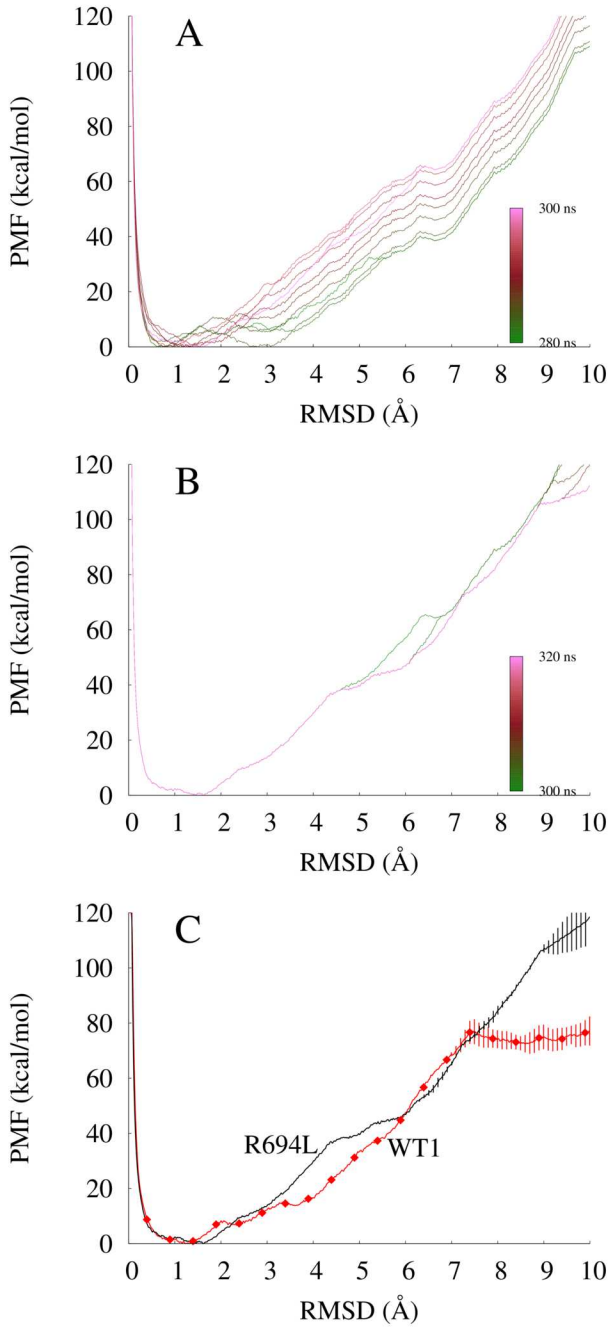

Figure 3.4: A, B) Convergence of cumulative PMFs for R694L peptide metadynamics. Conformational potential of mean force (PMF) in kcal/mol vs collective variable (RMSD of backbone compared to perfect helix) in Å during A) 280 to 300 ns and B) 300 to 320 ns. C) Final cumulative, conformational potential of mean force (PMF) in kcal/mol vs collective variable in Å from the R694L peptide metadynamics, in black. The PMF vs CV from the WT1 peptide metadynamics from our previous study is also shown here, in red diamonds [1]. Error bars represent standard deviations of the last 20 ns.

verged over the CV (Figure 3.4). The backbone RMSD of the peptide as compared to a perfect helix seemed an appropriate CV choice for a small (27-residue) membrane spanning α-helix because multiple refolding events were observed, as shown in Figure 3.3. During the entire metadynamics trajectory, the R694L peptide re-entered helical conformations during four separate events. Snapshots chosen from the trajectory are shown as insets in Figure 3.3. The cumulative PMF versus CV was averaged over 20 ns increments to assess convergence and the final PMF represented the average of the last 20 ns of cumulative PMFs. Figure 3.4 shows the convergence of the PMF during 280-300 ns and 300-320 ns. The cumulative PMF averaged from the last 20 ns for the WT and mutant peptides is shown in Figure 3.4 C. The R694L MSD PMF indicated a stable state with a low RMSD ($\leq$ 2 Å) and a metastable state with an RMSD in the range 4 to 6 Å. This is similar to the final PMF for the WT system, with a slight difference in the CV value of the metastable state. As expected, the R694L MSD did not have an appreciably altered stable state compared to the WT MSD and the R694L system was used for running equilibrium MD of the helical state.

To confirm that the stable state determined by metadynamics is the lowest free energy minimum for the R694L peptide, standard MD was continued from 300 ns of metadynamics and 304 ns of metadynamics, with starting CVs of 4.8 and 8.7 Å, respectively. Both of these trajectories rapidly refolded into stable α-helices in a few ns, as shown in Figure 3.5. Clearly, these trajectories did not start in equilibrium configurations and rapidly progressed through lower and lower energy configurations until reaching the lowest free energy well. This behavior was fully consistent with the PMF generated using metadynamics. Therefore, long NPT MD (300 ns) was next performed for the WT systems with the peptides in the energetically preferred α-helical state to allow determination of membrane and peptide equilibrium properties.

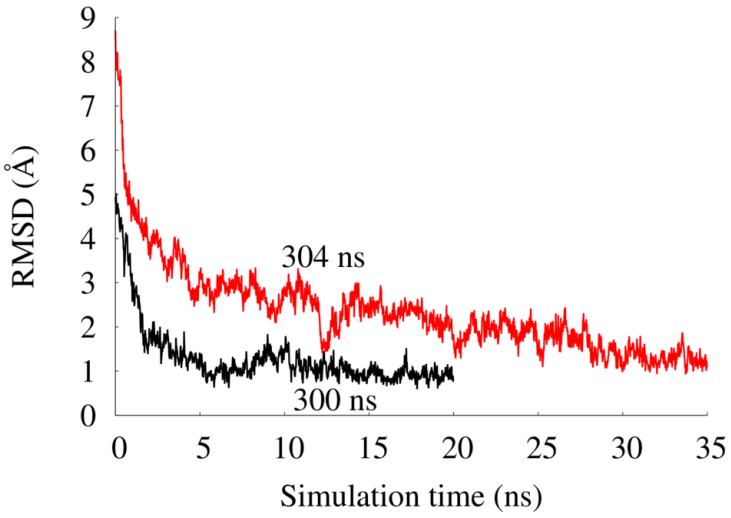

Figure 3.5: Standard molecular dynamics started from the metadynamics trajectory for R694L. Collective variable (RMSD of backbone compared to perfect helix) in Å vs simulation time in ns. MD started after 300 ns metadynamics in black and 304 ns metadynamics in red.

3.3.3 Model of HIV-1 gp41 MSD was stable as a membrane-spanning α-helix

The 300 ns production NPT MD simulations also support the conclusion that the α-helical state was stable for both the WT1 and R694L systems. In Figure 3.6 A and B, we show traces of the metadynamics CV observed during equilibrium MD launched from helical initial states for both systems, along with CV histograms. Both systems remained close to perfectly α-helical. The WT1 peptide seemed to have greater flexibility than the R694L peptide, as indicated by its wider RMSD distribution. In Figure 3.6 C, we show similar data for the WT1ΔChol system, which indicates that the WT peptide was also helical in a pure DPPC membrane. The absence or presence

of cholesterol seems to not have had an effect on the secondary structure of the MSD. Similar RMSD values were observed for the WT2, WT3, R694LR, WT2ΔChol, and WT3ΔChol systems (data not shown).

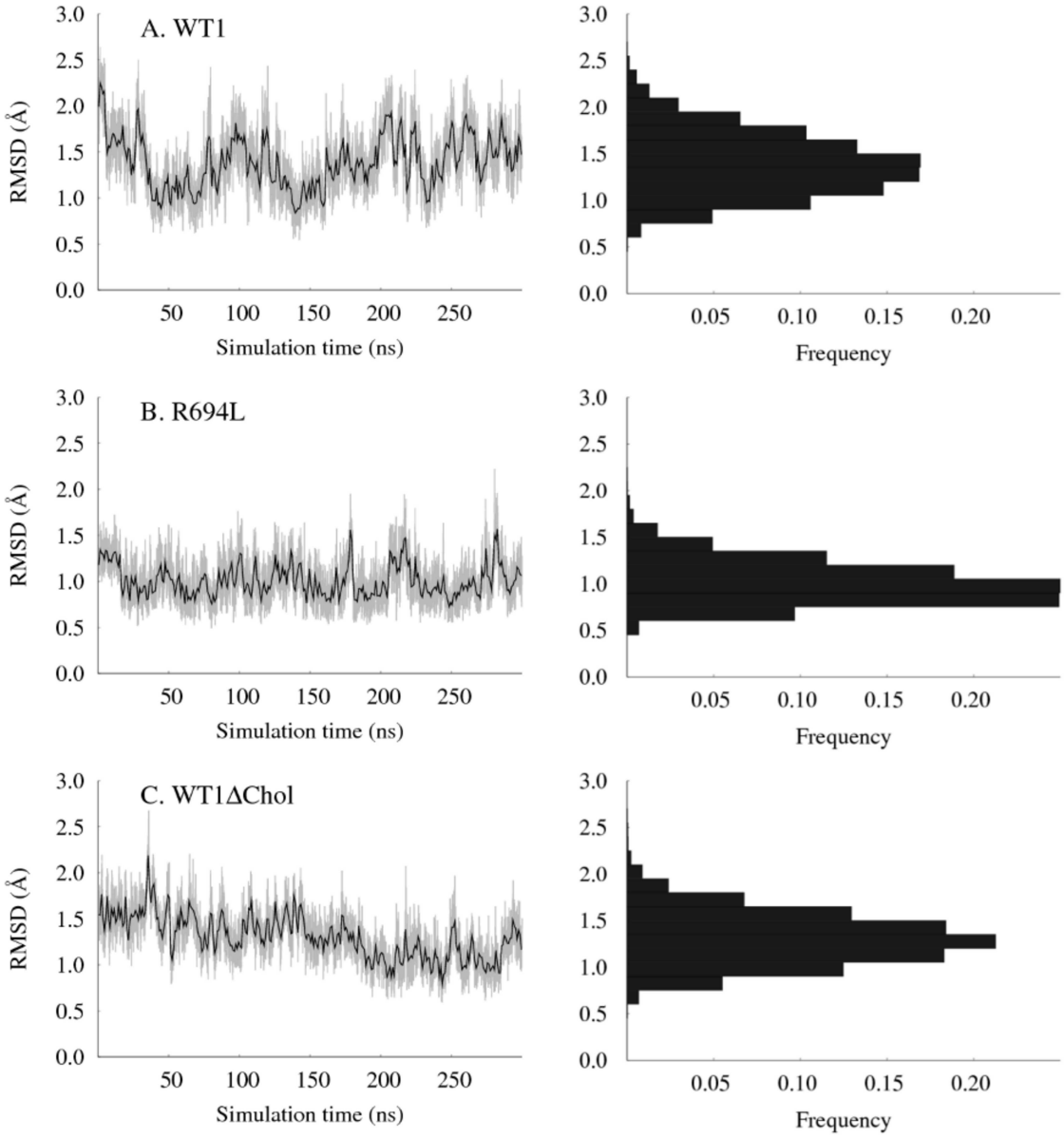

Figure 3.6: Collective variable (RMSD of backbone compared to perfect helix) in Å vs simulation time in ns and histogram of RMSD vs frequency for equilibrium MD of A) WT1, B) R694L, and C) WT1ΔChol systems.

Equilibrium MD further shows that these stable α-helices remained membrane-spanning. In Figure 3.7, we show mass-density profiles, averaged over the entire production phase, for various components in the WT (averaged over 3 replicas), R694L, and WTΔChol (averaged over 3 replicas) systems. The right-hand panels in this figure show global density profiles that establish the z-span (membrane normal) of the bilayer, and the accompanying left-hand panels show the scaled local density of protein and water within 4 Å of the protein. Clearly all proteins spanned the bilayer. Also highlighted in the left-hand panels of Figure 3.7 are the mass density distribution of atoms in the 694 residue; we see that R694 density overlaps with water density within 4 Å, indicating the R694 was solvated in both the WT and WTΔChol systems. R694L was not solvated, as expected. Mass density profiles for other systems are in Appendix Figure 3.17.

3.3.4 Midspan arginine of MSD was solvated by a robust water defect

All systems with the wild-type sequence showed a solvated midspan arginine and this was further quantified by finding the number of water molecules (within 4 Å of the protein) uniquely associated with each residue, as seen in Figure 3.8. Panel A shows the number of unique water molecules per residue for the WT1, WT2, and WT3 systems. The N- and C- termini were highly solvated, as they interacted with the lipid headgroup region. The midspan arginine (local residue 14) interacted with 4.8 ± 0.03 (SE) water molecules in the WT1, WT2, and WT3 systems. This resulted in solvation along the same side of the helices as the midspan arginine, which corresponds in the figure to the solvation in the C-terminus every few residues (local residues 17, 21, 25). As expected, the R694L system did not have any water molecules that solvated the R694L residue, as seen in Panel B. After the setup of the R694L system with mutation of residue 694 to leucine (see System Setup), the water molecules surrounding residues

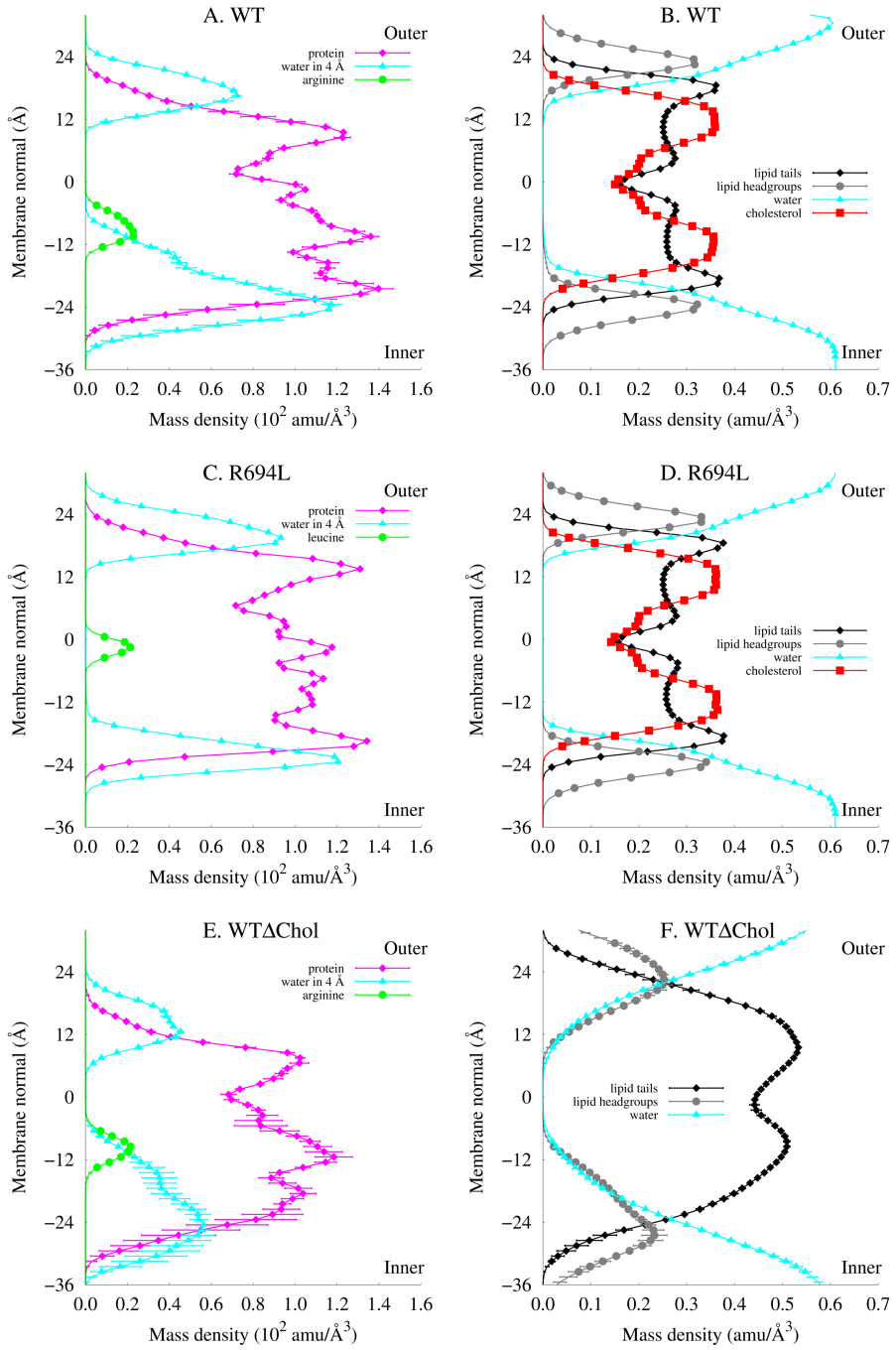

Figure 3.7: A, C, E) Mass density in 10^2 amu/Å^3 along membrane normal in Å during equilibrium MD for various components of the system: protein, local water (defined as within 4 Å of protein), and midspan residue (arginine or leucine). B, D, F) Mass density in amu/Å^3 along membrane normal in Å during equilibrium MD for various components of the system: lipid tails, lipid headgroups, water, and cholesterol. A) and B) are averages for WT1, WT2, and WT3 systems. E) and F) are averages for WT1ΔChol, WT2ΔChol, and WT3ΔChol systems. All statistics are from the last 100 ns of each trajectory and error bars in A, B, E and F represent standard error.

R694L to L700 took a mere 6 ns or less to vacate the membrane. The water molecules did not return to these residues during 300 ns MD. To check initial condition bias of the midspan arginine-containing systems, another system (R694LR) was created from the R694L system where the midspan leucine was spontaneously mutated back to the WT sequence and simulated with MD. In R694LR, the first water molecule interacted with the midspan arginine within 9 ns. The water profile for R694LR resembled the WT1, WT2, and WT3 systems within 100 ns. This simulation confirmed that the WT MSD can cause influx of water into the membrane during ns timescales and that solvation of the midspan arginine is not a result of the initial condition from pulling the peptide into the membrane. In Panel C, the cholesterol-free systems were similarly solvated, with an average of 6.7 ± 0.44 (SE) water molecules uniquely associated with the midspan arginine. Evidently midspan arginine solvation does not require cholesterol and this so-called "water defect" is a robust feature of the MSD.

The midspan arginine has 5 hydrogens available to hydrogen bond with water molecules and lipid headgroups. The actual number of hydrogen bonds a transmembrane arginine forms in bilayers depends on lipid type and arginine bilayer depth, but it ranges from 3 to 5 [61, 96, 106]. Figure 3.9 A and B show that the midspan arginine engaged in 3 to 4 hydrogen bonds throughout the trajectories in both the cholesterol-containing and cholesterol-free systems. Figure 3.9 C and D show the number of hydrogen bonds from R694 to water in the systems. Hydrogen bonds to water represented most of the hydrogen bonds; the remainder are to lipid phosphate oxygens, lipid carboxyl oxygens, and (for the cholesterol-containing systems), cholesterol hydroxyl oxygens (data not shown).

Transmembrane helices that have a charged residue that "snorkels" to water and lipid headgroups often tilt to facilitate the solvation. This was indeed the case with the MSD in both cholesterol-containing and cholesterol-free bilayers, as seen in Fig-

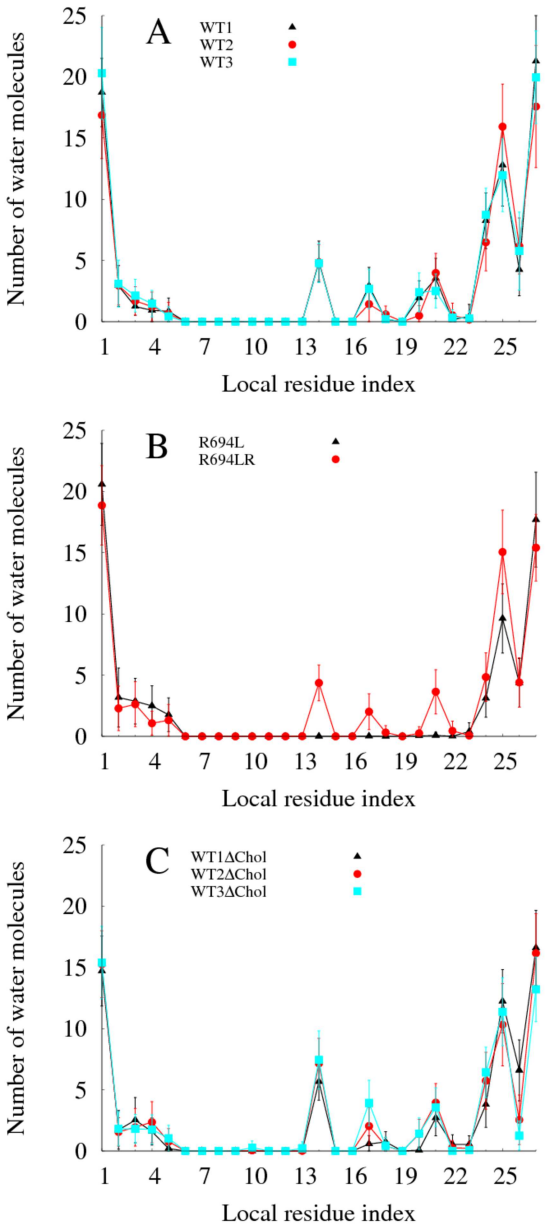

Figure 3.8: Number of water molecules within 4 Å of protein vs amino acid that each water molecule is uniquely attributed to during equilibrium MD. All statistics are from the last 100 ns of each trajectory and error bars represent standard deviation.

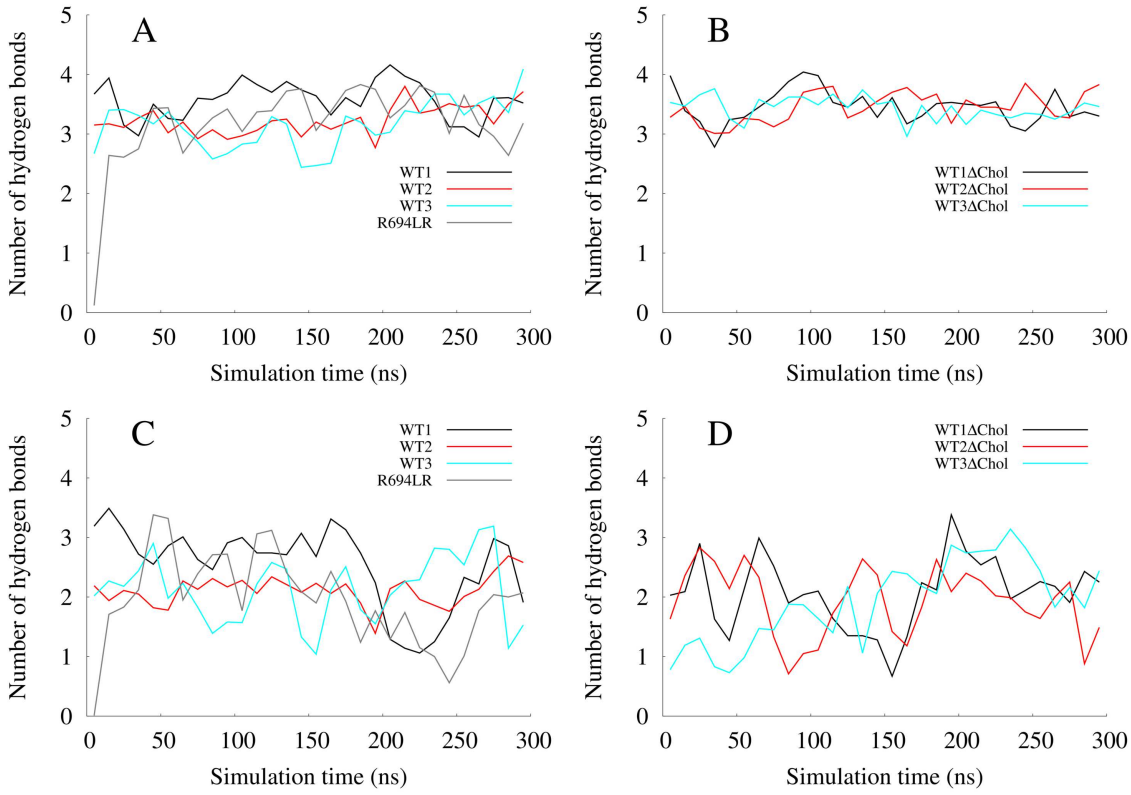

Figure 3.9: Number of total hydrogen bonds donated from R694 vs simulation time in ns for A) cholesterol-containing systems WT1, WT2, WT3, and R694LR and B) cholesterol-free systems WT1ΔChol, WT2ΔChol, and WT3ΔChol. Number of hydrogen bonds donated from R694 to water vs simulation time in ns for C) cholesterol-containing systems WT1, WT2, WT3, and R694LR and D) cholesterol-free systems WT1ΔChol, WT2ΔChol, and WT3ΔChol

ure 3.10. In Panel A, the WT systems in cholesterol-containing membranes had similar tilts from membrane normal during the last 100 ns that fluctuate around 19.5° ± 1.4 (SE). In Panel B, the tilt of the helix in the R694LR system was similar. However, the R694L systems sampled an average of 14.3° ± 3.1 (SD) from membrane normal. The midspan leucine did not "snorkel" and therefore the R694L had a lower tilt than the WT1, WT2, and WT3 systems. The WT systems in cholesterol-free bilayers, shown in Panel C, also tilted but to widely variable degrees: averages of 7.2° ± 3.0, 22.6° ± 3.0, 39.5° ± 3.7 (SD) for WT1ΔChol, WT2ΔChol, WT3ΔChol, respectively. This suggests that the cholesterol-containing and cholesterol-free bilayers

accommodated the “water defects” differently.

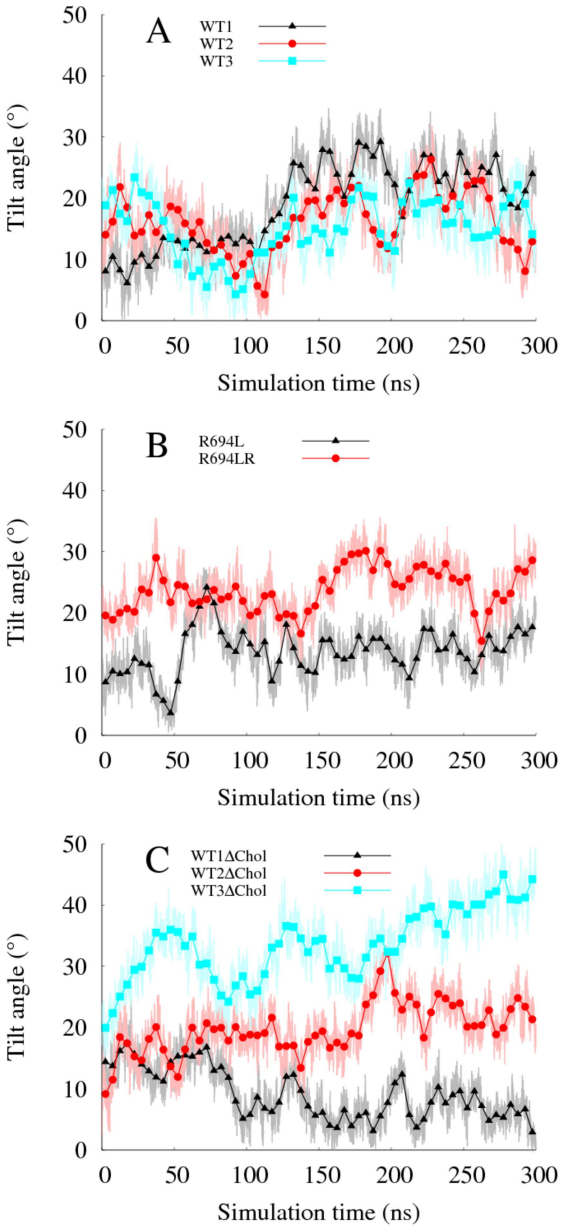

Figure 3.10: Tilt angle in degrees vs simulation time in ns.

3.3.5 Localized thinning near water defect was dependent on cholesterol

The comparison of membrane thickness in intervals of 100 ns is useful for assessment of membrane stability and for evaluation of any changes in the membrane environment near the peptide. Shown in Figure 3.11 are 2D maps of membrane thickness and standard deviation in 100 ns intervals for WT1, R694L, WT1ΔChol, and WT3ΔChol systems. The maps are centered on the location of the peptide, of which the backbone is shown in black dots overlaid on the maps. The midspan arginine (or leucine) nonhydrogen sidechain is shown in white dots. The cholesterol-containing membranes were generally thicker than the cholesterol-free membranes here, as cholesterol has been shown to order the lipid tails in the liquid phase, which increases the thickness [16, 107]. The membrane thickness and fluctuations for WT1 were consistent across the 3 time intervals and were uniform over the total area, except near the peptide. The WT1 membrane was thinner near the peptide and water defect and had greater fluctuations there. The R694L system had uniform thickness and fluctuations, even at the location of the peptide. Therefore, the membrane thinning in the WT1 system was due to the midspan arginine and water defect specifically, and not the presence of the peptide. In the cholesterol-free systems shown (WT1ΔChol and WT3ΔChol), membrane thickness and fluctuations were not uniform and the peptides were located in areas of different thicknesses. In WT1ΔChol, the peptide was located in a thicker section of the membrane, but the thickness was fluctuating there. In WT3ΔChol, the peptide was in a thinner section of the membrane and the thickness was also fluctuating there. In the cholesterol-free membranes, the thickness of the location of the peptide correlated to its tilt angle; the thinner the membrane section, the greater the tilt. It seemed that in the cholesterol-containing membranes, the water defect influenced the local thickness but in the cholesterol-free membranes, the water defect did not influence the local thickness. This helps explain why the

peptide tilts were consistent in the cholesterol-containing membranes with a water defect, but the tilts in the cholesterol-free membranes with a water defect depended on the thickness at the exact location of the peptide. This relationship can be seen in Figure 3.12. However, the water defect did cause thickness fluctuations in both cholesterol and cholesterol-free membranes, which functioned to solvate the midspan arginine also.

The cholesterol-containing membranes did not change appreciably in thickness between intervals. The initial membrane configuration seemed to be stable on a 100 ns timescale. The cholesterol-free bilayers did change from the first 100 ns interval to the last 2 intervals, as it took them over 100 ns to equilibrate (see Figure 3.2). Maps of membrane thickness and standard deviation for the other systems are shown in Appendix Figure 3.18. Importantly, the membrane thickness map for the protein-free DPPC/Chol system was completely uniform, with an average of 48 Å. This is consistent with both experimental and computational studies [16, 107].

We have shown that membrane thickness fluctuations helped solvate the water defect. Next, we look at actual water movement in the inner leaflet. In Figure 3.13 are 2D maps of average minimum distance of water from the bilayer interface in the inner leaflet and standard deviation for WT1, WT1ΔChol, and WT3ΔChol systems. The WT1 system had uniform average minimum water location from the membrane interface, except at the water defect. Water molecules came to within 14 Å of the WT1 bilayer middle there and the membrane thickness there is 37.5 Å. The cholesterol-free systems had the closest water molecules in thin areas of the membrane and at the water defect. For WT1ΔChol, the protein was in an area with 43 Å thickness and water molecules came to within 11.6 Å. For WT3ΔChol, the protein was in an area with 32.5 Å thickness and water molecules came to within 11.8 Å. The water distance was also highly fluctuating at the water defect for all three systems. It seems that

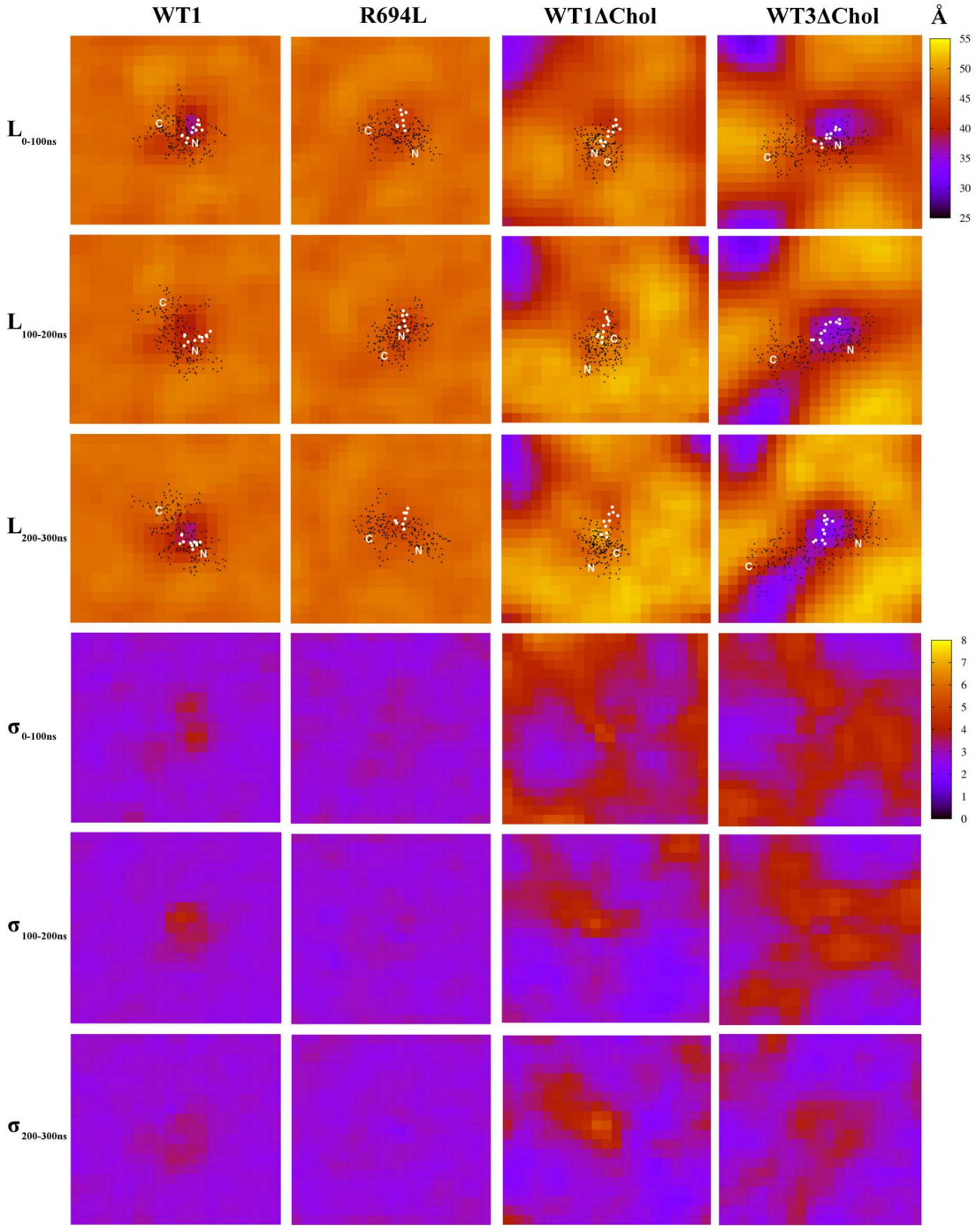

Figure 3.11: Maps of membrane thickness, L, (top) in Å and standard deviation, σ_{L}, (bottom) in Å for the WT1, R694L, WT1ΔChol, and WT3ΔChol systems during 100 ns intervals of equilibrium MD. Overlaid on the maps are the x and y positions of the non-hydrogen atoms of the peptide (black circles) and the 694 residue (white circles) from the last frame of the trajectories. The N- and C-termini are labeled.

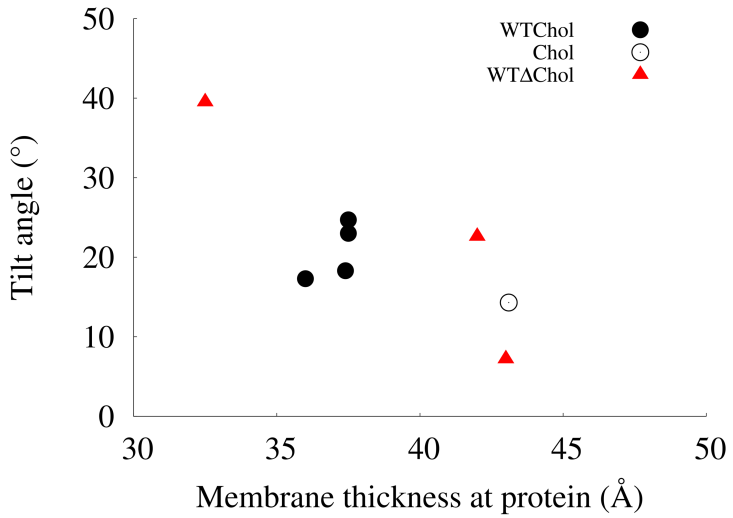

Figure 3.12: Average tilt angle in degrees from last 100 ns vs membrane thickness at protein in Å for WT peptides in cholesterol in black closed circles (WT1, WT2, WT3, R694LR), mutant peptide in cholesterol in black open circle (R694L), and WT peptides in cholesterol-free membranes in red triangles (WT1ΔChol, WT2ΔChol and WT3ΔChol).

water and lipid headgroups worked in unison to accommodate the midspan arginine. Maps of average minimum water distance along membrane normal in the inner leaflet for the other systems are shown in Appendix Figure 3.19.

3.3.6 Summary and conclusions of Aim I

A model of HIV-1 gp41 monomeric MSD in a model viral bilayer was shown to be stable as an α-helical peptide in a membrane spanning configuration for 300 ns. The highly-conserved midspan arginine was solvated due to water and membrane thickness fluctuations in both cholesterol-containing and cholesterol-free bilayers. The water defect did not depend on local lipid distribution or system initial condition. However, only in bilayers with cholesterol was the helix tilt unambiguously associated with the water defect. In the cholesterol-free bilayers, the helix tilt also depended on the membrane thickness. In the cholesterol-containing bilayers, membrane thinning was localized at the water defect; without solvation of a midspan residue, there was no

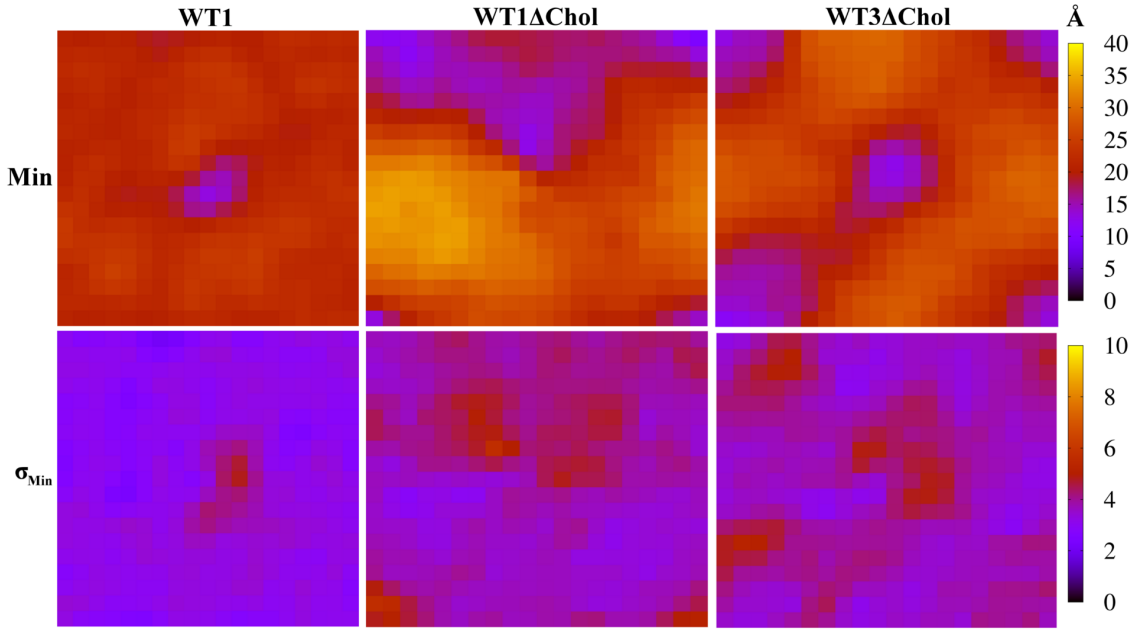

Figure 3.13: Maps of average minimum distance along membrane normal in the inner leaflet of water molecules from global center of mass of lipid bilayer, Min (top), in Å and standard deviation, σ_{Min}, (bottom) in Å for WT1, WT1ΔChol, and WT3ΔChol systems.

membrane thinning at the peptide. A consistent helix tilt may be important for a protein that trimerizes, even though it is not known if or when the MSD exists as a trimer. We can also speculate on the importance of a water defect at the MSD. Since gp41 is a metastable protein whose conformational rearrangements into a more stable state are thought to be the driving force for HIV-1 fusion/infection, perhaps the water defect represents a local membrane metastability that also helps drive the fusion process.

3.3.7 Appendix

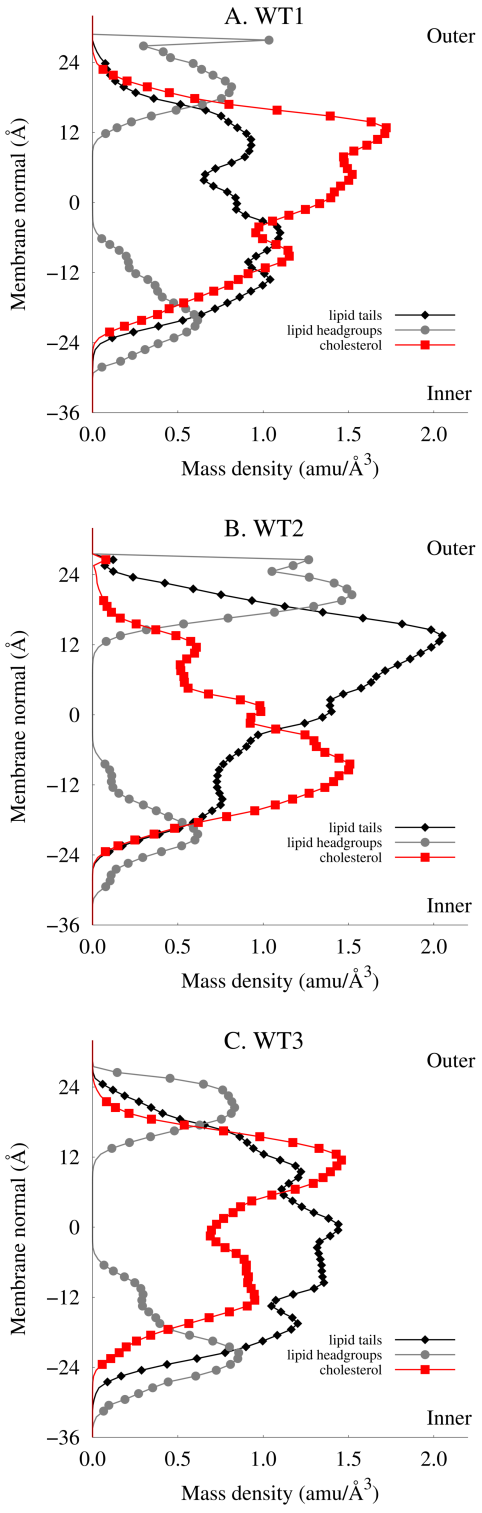

Figure 3.14: Local mass density (within 6 Å of protein) in amu/Å^3 along membrane normal in Å during equilibrium MD of lipid tails, lipid headgroups, and cholesterol for A) WT1, B) WT2, and C) WT3 systems. All statistics are from the full 300 ns of each trajectory.

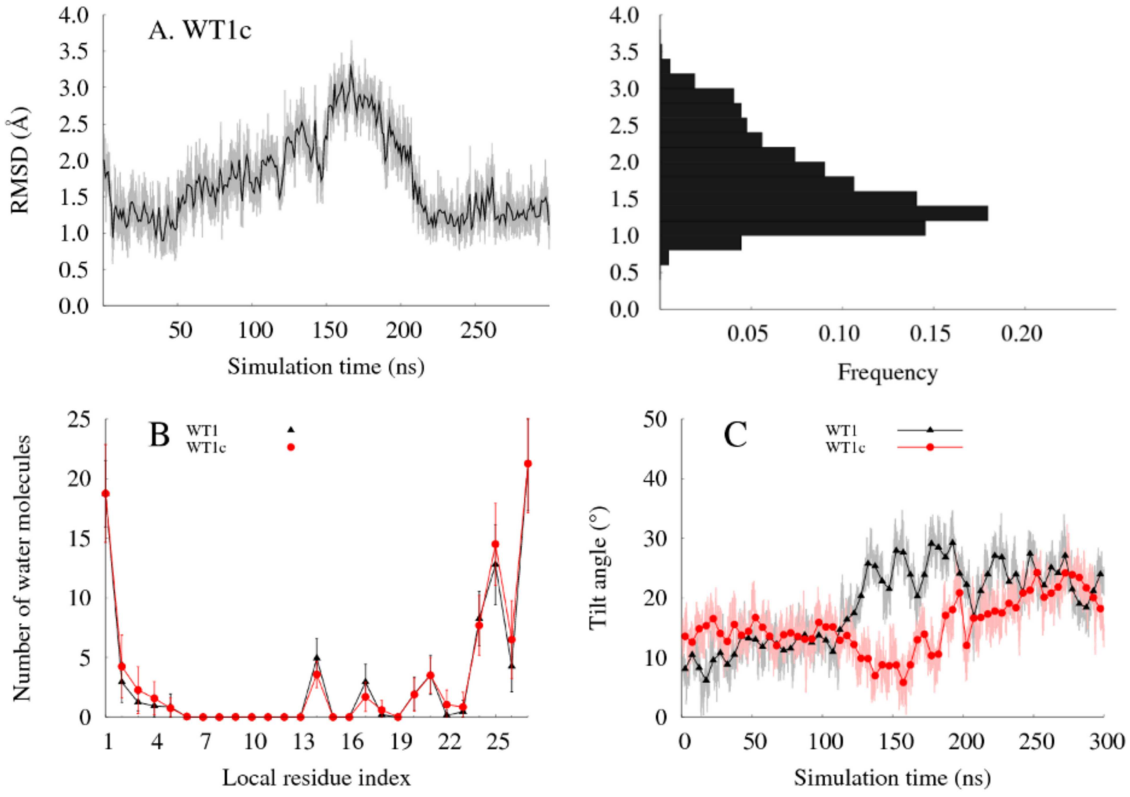

Figure 3.15: Comparison of WT systems run with more aggressive (WT1) and less aggressive (WT1c) MD parameters (see Methods). A) Collective variable (RMSD of backbone compared to perfect helix) in Å vs simulation time in ns and histogram of RMSD vs frequency for equilibrium MD of WT1c. B) Number of water molecules within 4 Å of protein vs amino acid that each water molecule is uniquely attributed to during equilibrium MD for WT1 and WT1c systems. All statistics are from the last 100 ns of each trajectory and error bars represent standard deviation. C) Tilt angle in degrees vs simulation time in ns for WT1 and WT1c systems.

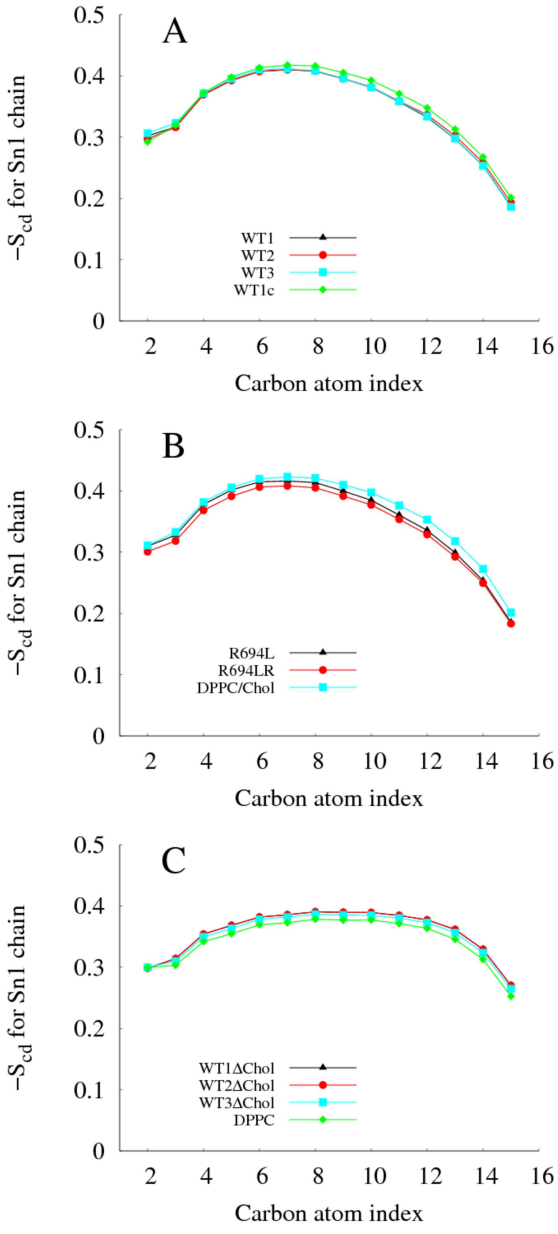

Figure 3.16: Sn1 order parameters vs carbon atom index from last 100 ns of trajectory.

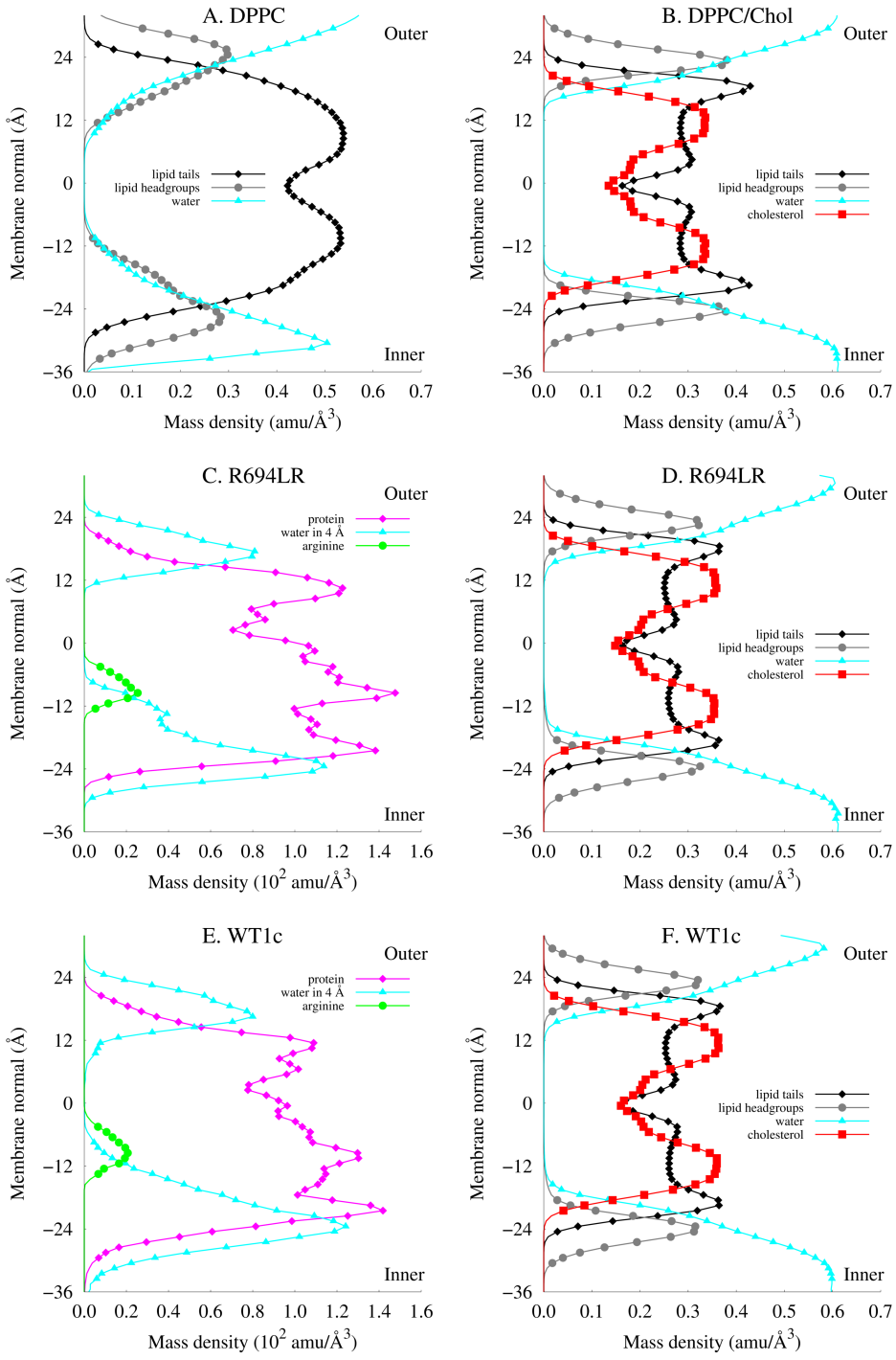

Figure 3.17: A, B, D, F) Global mass density in amu/Å^3 along membrane normal in Å during equilibrium MD for various components: lipid tails, lipid headgroups, water, and cholesterol for DPPC, DPPC/Chol, R694LR, and WT1c systems. C, E) Local mass density in 10^2 amu/Å^3 along membrane normal in Å during equilibrium MD for various components: protein, local water (defined as within 4 Å of protein), and midspan arginine for R694LR and WT1c systems. All statistics are from the last 100 ns of each trajectory.

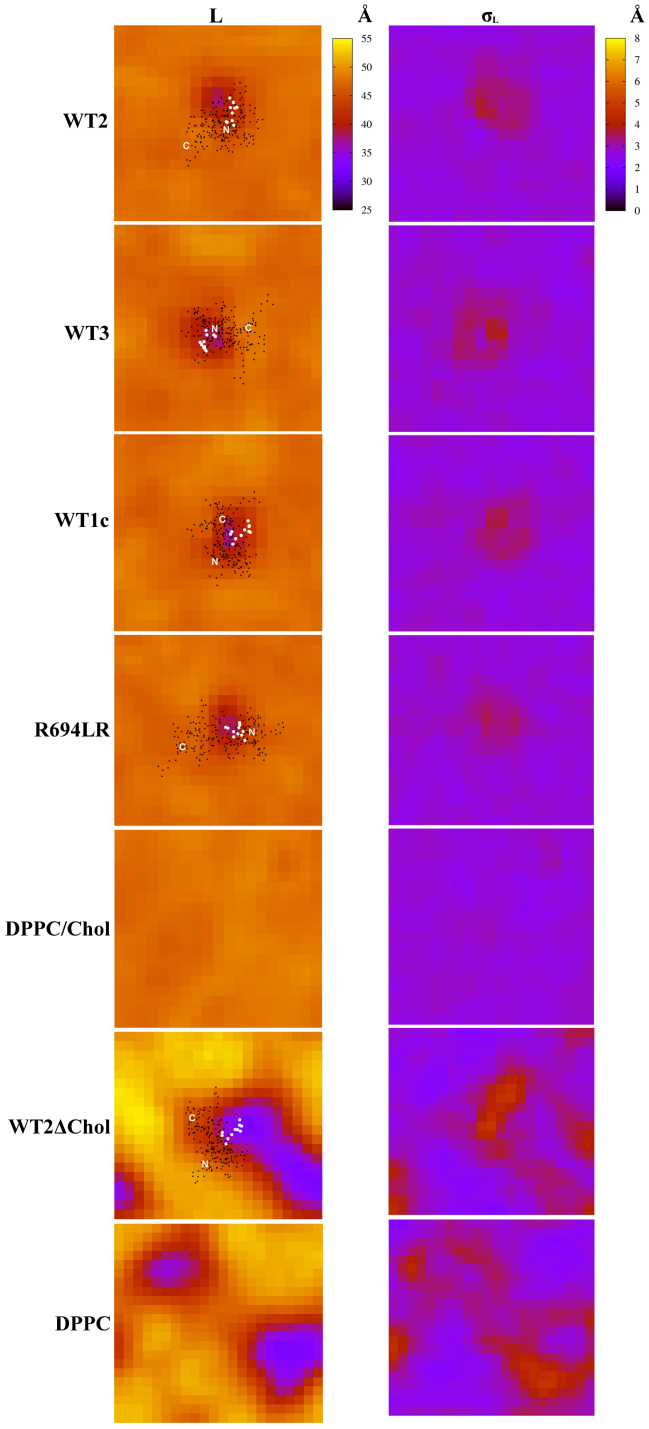

Figure 3.18: 2D maps of membrane thickness, L, (left column) in Å and standard deviation, σ_{L}, (right column) in Å for the WT2, WT3, WT1c, R694LR, DPPC/Chol, WT2ΔChol, and DPPC systems from equilibrium MD. Overlaid on the maps are the x and y positions of the non-hydrogen atoms of the peptide (black circles) and the 694 residue (white circles) from the last frame of the trajectory. The N- and C- termini are labeled. Statistics are from the last 100 ns of each trajectory.

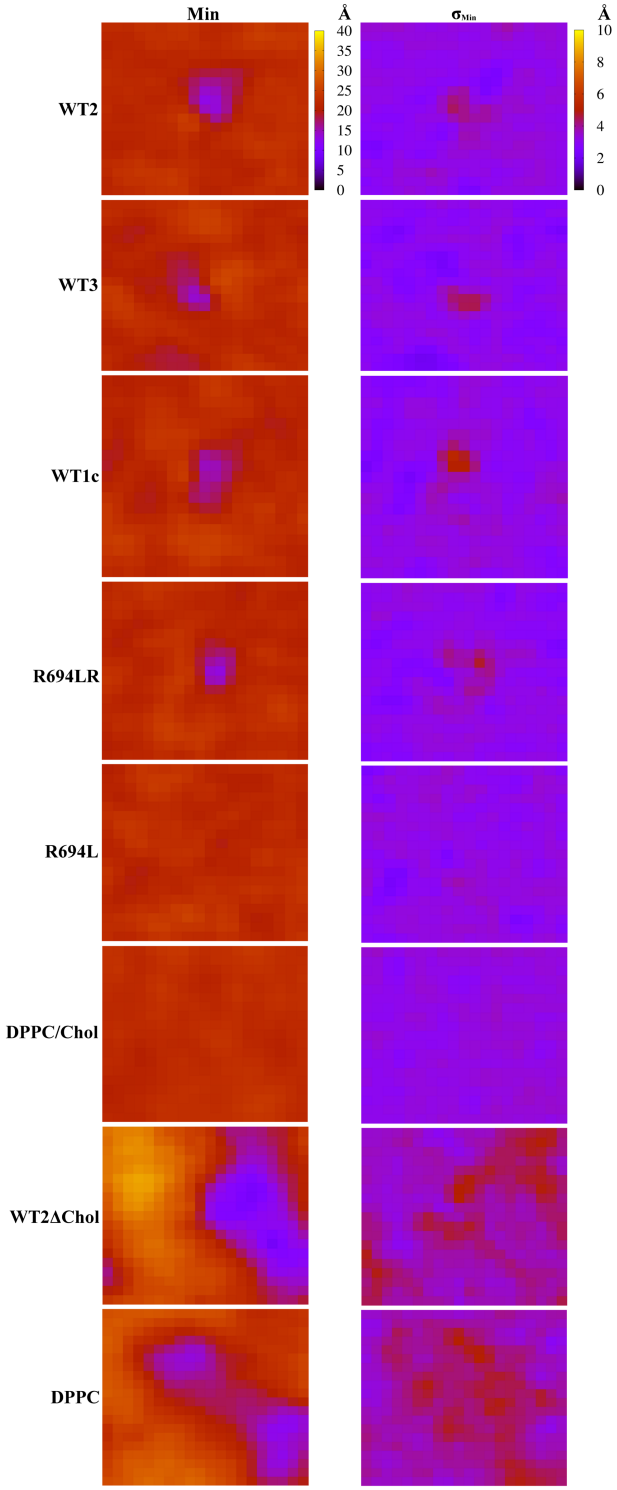

Figure 3.19: 2D maps of average minimum distance of water molecules along membrane normal in inner leaflet from global center of mass of lipid bilayer, Min (left column), in Å and standard deviation, σ_{Min}, (right column) in Å for WT2, WT3, WT1c, R694LR, R694L, DPPC/Chol, WT2ΔChol, and DPPC systems. All statistics are from the last 100 ns of each trajectory.

4. Aim II: Dynamics of HIV-1 model viral membranes with cholesterol

4.1 Introduction to Aim II

A common model membrane for HIV-1 research is 50/50 DPPC/Chol and this composition has been studied both experimentally and computationally. Computational studies of membranes with 50% cholesterol are strongly influenced by the initial coordinates of the lipids, because diffusion of this composition occurs on longer timescales than previously computationally accessible [108, 109]. Therefore, less is known about the dynamics and configurations of bilayers relevant for HIV-1 than other model membranes with much less cholesterol. In Aim I, three replicas (WT1, WT2, and WT3) were used to represent different local lipid distributions around the MSD, since sufficient diffusion around the protein was beyond the 300 ns timescale of Aim I. Another way of sampling phase space besides multiple replicas is to run extremely long MD (infinitely, in theory) to generate an approximation of the ensemble average, according to the ergodic principle. Since HIV-1 fusion and infection are strongly influenced by lipid composition, Aim II simulated WT1, WT2, and R694L from Aim I for 6.45 to 9.98 μs, an order of magnitude greater than previous MD, using the highly specialized, state-of-the-art computer Anton [110, 111] to explore ergodicity of the membrane and study the dynamics of monomeric MSD with lipids and cholesterol molecules in a model viral bilayer. We aimed to answer the following questions:

1. Is the diffusion of lipids and cholesterol during 10 μs MD consistent with experiments?

2. Is the water defect stable on this timescale?

3. Is this timescale long enough for the membrane to "forget" its initial state?

4.2 Methods

4.2.1 μs trajectories on Anton

Systems WT1, WT2, and R694L after 300 ns NPT MD from Aim I were further simulated for 9.98, 6.45, and 8.06 μs NPT MD, respectively, on the Anton supercomputer [110, 111]. As before, the CHARMM force field with recent lipid-based corrections and explicit TIP3P water were used [75–78]. Verlet integration was applied every 2 fs and long-range electrostatics were handled with the Gaussian Split Ewald method [112]. The pressure, set to 1 atm, was handled semi-isotropically with the Martyna-Tobias-Klein barostat and the Nosé-Hoover thermostat was set to 310 K [80, 113]. Trajectories were visualized with VMD [74]. See Appendix C for the complete list of configuration parameters.

Trajectories from Anton were only available with all coordinates wrapped into the primary periodic cell. The trajectories had to be postprocessed into unwrapped versions in order to compute mean-squared displacements. Unwrapping of all molecules from one frame to the next to determine the "true" displacement requires that the displacement of each molecule is smaller than half the periodic box (in each cartesian direction). In order for unwrapping to generate true displacements and to avoid intra-leaflet fictitious forces, center-of-mass motion was not removed during the simulations. The thermostat available in the Anton software was the Nosé-Hoover, which rescales velocities. This type of thermostat with no removal of center-of-mass motion resulted in cooling of WT1, WT2, and R694L systems to 306.7, 302.6, and 302.2 K, respectively. This cooling is shown in Appendix Figure 4.9 and is related to the flying ice cube effect [114–116]. However, the temperatures stabilized in our simulations instead of continuously cooling.

4.2.2 Calculations of observables

Lipid order parameters, tilt of helix with respect to membrane normal, unique water molecules per residue, 2D membrane thickness maps, and local density were calculated the same way as in Chapter 3. For calculation of diffusion coefficients, the trajectories were rewrapped in x and y around the leaflet center of mass and then unwrapped in x and y. The mean-squared displacements (MSD) of lipids, cholesterol molecules, and protein were measured over all time origins for each leaflet independently. Using Einstein's equation,

$$D_{lat} = \lim_{t\to\infty} \frac{1}{4Nt} \sum_{i=1}^{N} \langle [r_i(t) - r_i(0)]^2 \rangle, \tag{4.1}$$

the lateral diffusion coefficients were calculated when the MSD was linear with time (for time intervals between 1 and 5 μs for DPPC/cholesterol or between 0.1 and 2.5 μs for peptide). Graphs of the mean squared angular displacements (MSAD) vs time were used to similarly calculate the 1D rotational diffusion coefficients. The tilt of cholesterol molecules was calculated by determining the angle between membrane normal and a vector connecting carbons C3 and C17. The location of either cholesterol hydroxyl oxygen or DPPC phosphorous atoms in 4 x 4 Å^2 sections of the peptide-centered trajectories were used to create maps of average mass in amu. The radial distribution function was calculated using the *g(r)* plugin in VMD and a δ of 0.1 Å [117]. The pair selections chosen were the cholesterol hydroxyl oxygen O3 with respect to itself, DPPC phosphate oxygens O11, O12, O13, O14 to cholesterol hydroxyl hydrogen H3', and DPPC carbonyl oxygen O22 or O32 to cholesterol hydroxyl hydrogen H3'. See Appendix Figure 4.10 for a picture of each lipid component with important atoms highlighted.

4.3 Aim II results

4.3.1 System properties were consistent between ns and μs MD

System properties measured from 6.45 to 9.98 μs MD for WT1, WT2, and R694L systems were consistent with measurements from 300 ns MD in Aim I. The WT and mutant peptides remained membrane spanning for the duration of the trajectories. The peptides were α-helical and stable, as seen in plots of RMSD in Appendix Figure 4.11. The midspan arginines of WT1 and WT2 remained solvated, while the mutant did not have a water defect, as seen in Appendix Figure 4.12. Also consistent with Aim I, the MSD peptides had a tilt angle associated with either the presence (WT1 and WT2) or absence of a water defect (R694L), as seen in Appendix Figure 4.12.

4.3.2 Local membrane distribution changed slowly

Next we examined initial condition bias of protein-embedded membrane simulations and the degree to which erogodicity is achieved in these simulations on the multiple microsecond timescale. Figures 4.1 and 4.2 show snapshots along membrane normal of the lower leaflets of the three systems at the beginning and end of the trajectories. Either lipids or cholesterol molecules within 6 Å of the protein at the beginning of the simulation are colored uniquely in the peptide-centered snapshots. The locations of these same membrane components are also highlighted at the end of the simulation and show that membrane components initially local to the peptides mostly diffused away on the microsecond timescale. However, these snapshots also indicate that the lipid and cholesterol molecules in the WT2 and R694L systems did not move as far as in the WT1 system, as some of the molecules still remain in the vicinity of the peptide.

The mass densities of lipid and cholesterol molecules local to the peptide quantita-

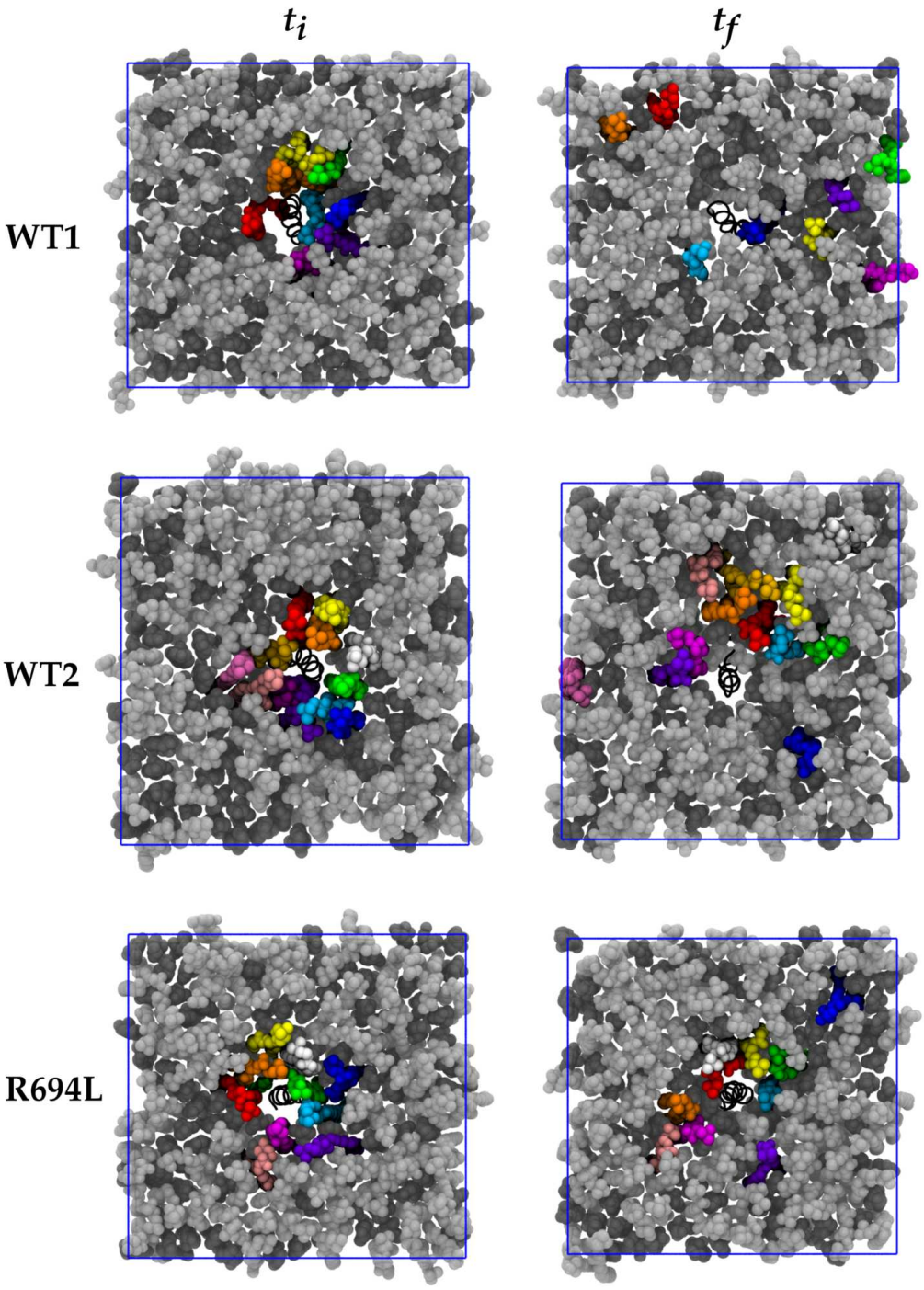

Figure 4.1: DPPC lipid molecules within 6 Å of protein in the lower leaflet in bright colors at the first simulation frame (t_i) and the same molecules at the final simulation frame (t_f). Lipid molecules are in light gray and cholesterol in dark gray. The protein is represented as a black spiral. The blue square represents the x and y periodic boundary conditions.

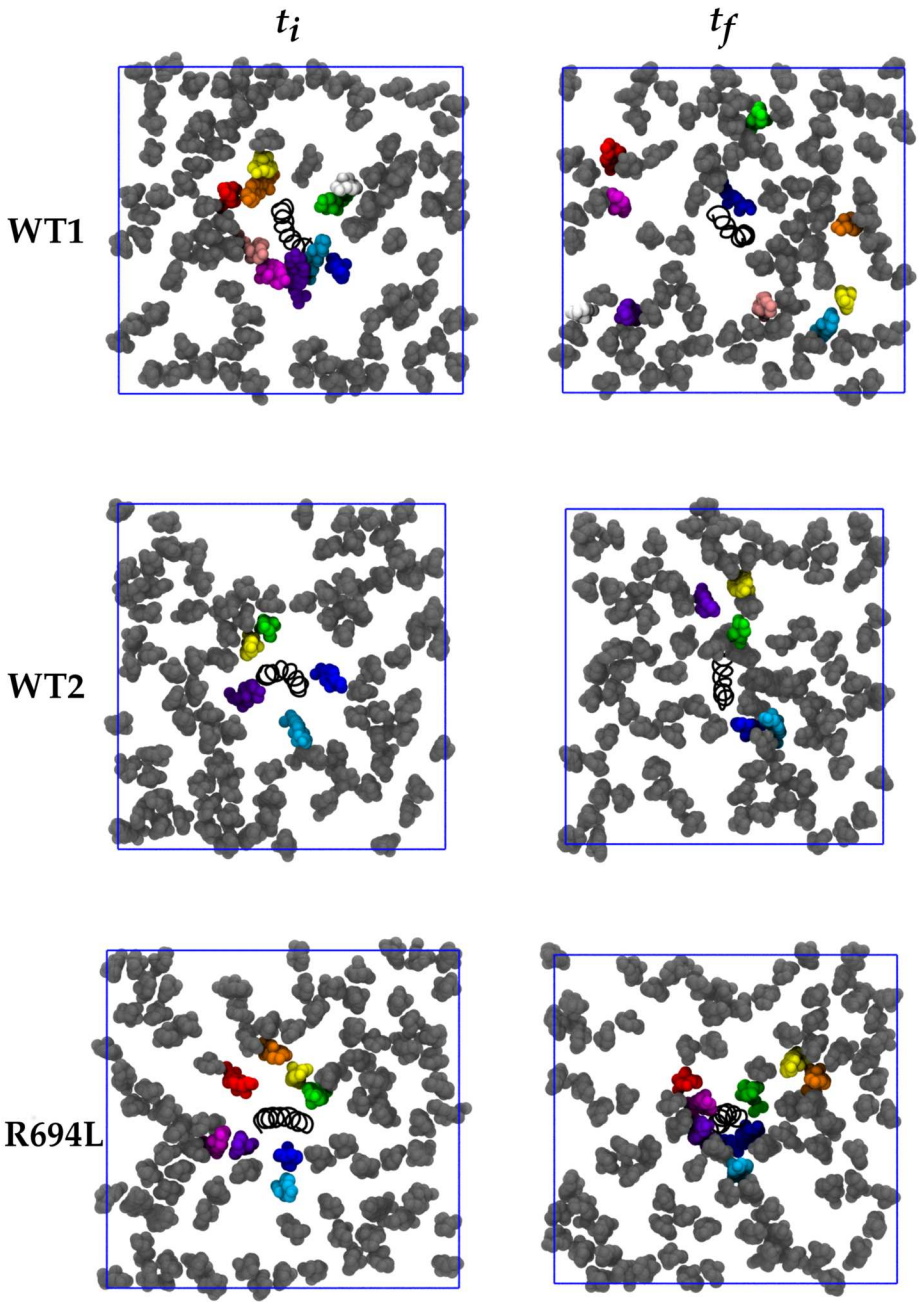

Figure 4.2: Cholesterol molecules within 6 Å of protein in the lower leaflet in bright colors at the first simulation frame (t_i) and the same molecules at the final simulation frame (t_f). Cholesterol molecules are in dark gray and the protein is represented as a black spiral. Lipid molecules are not shown for clarity. The blue square represents the x and y periodic boundary conditions.

tively assess ergodicity in the systems and are shown in Figure 4.3. The local density profiles of the first two μs are different for each system, reflecting the choice in Aim I to create systems with different initial conditions. Complete ergodicity would be achieved if WT1 and WT2, which are replicas, have the same local density at the end of the simulations. The density profiles of the last two μs are similar, and the peptides in WT1 and WT2 have greater local cholesterol than local DPPC density in the lower leaflet. The profiles are not equal, however, and these trajectories are evidently not long enough for equilibration of lipid distribution.

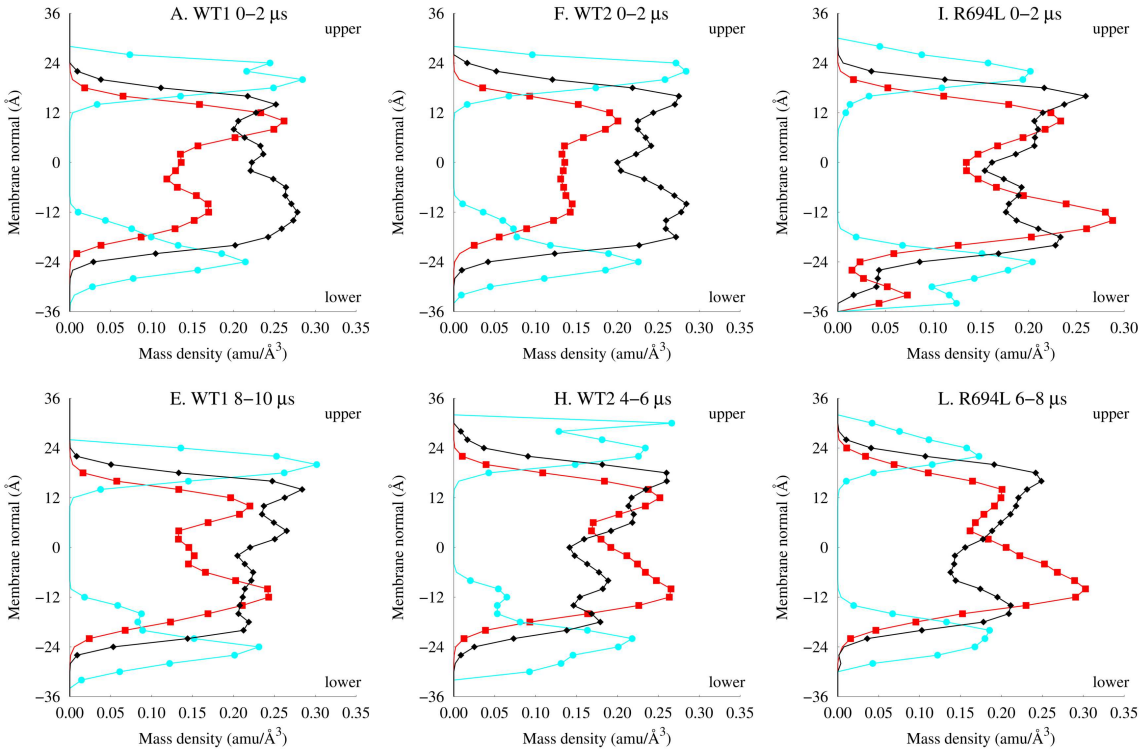

Figure 4.3: Mass density in amu/Å^3 within 8 Å from protein along membrane normal for cholesterol (red squares), lipid headgroups (cyan circles), and lipid tails (black diamonds) for A, B, C) first 2 μs MD and D, E, F) last 2 μs MD for WT1, WT2, and R694L systems.

The trajectories did not show any clustering between membrane components,

which remained randomly distributed. This was determined by radial distribution function, *g(r)*, over time and averaged for each system. The *g(r)* for specific DPPC oxygen atoms with respect to cholesterol hydroxyl hydrogen atoms or cholesterol hydroxyl oxygen atoms with respect to itself were identical between systems, as seen in Appendix Figure 4.13, and are consistent with similar, previous studies [107].

The locations of bilayer components that interact with the MSD for many μs in the lower leaflet are shown by 2D maps of either average cholesterol hydroxyl oxygen or DPPC phosphorous mass in amu per frame for the 3 systems in Figure 4.4. WT1 had uniform distribution of cholesterol but increased DPPC mass near the protein, while WT2 and R694L had increased cholesterol density. The bilayers were able to better configurationally sample on the μs timescale, but there were still lipid-peptide or cholesterol-peptide interactions that persisted for the majority of the length of the simulations.

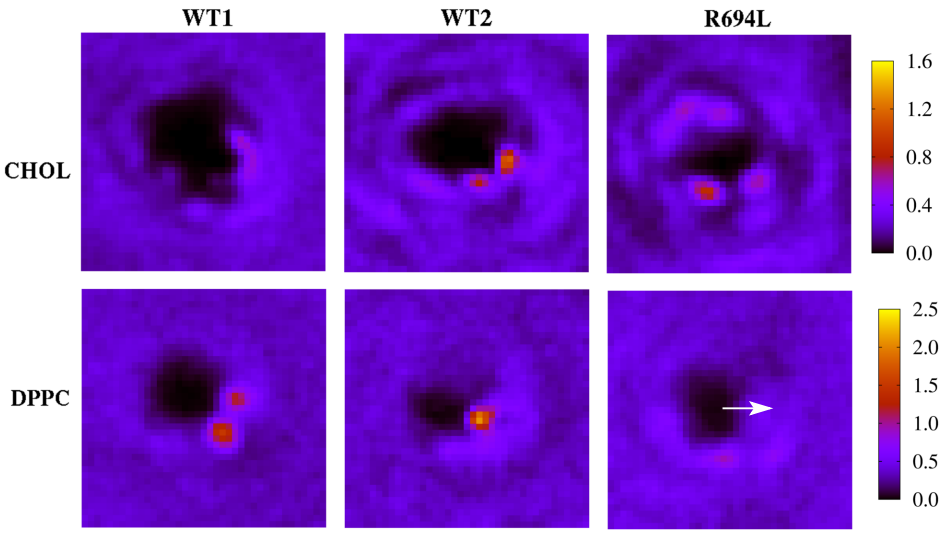

Figure 4.4: Maps of average mass (amu) per frame 20 Å from protein in the lower leaflet for WT1, WT2, and R694L. First row, cholesterol hydroxyl oxygen mass and second row, DPPC phosphorous mass. The peptides are centered in x and y and the azimuthal orientation of the midspan residue (arginine for WT1 and WT2 and leucine for R694L) is aligned along the x-axis in each map, as shown by the white arrow in the bottom right box.

4.3.3 Translational and rotational diffusion of HIV-1 model bilayer comparable to experiments

The 6.45 to 9.98 μs simulations allowed sufficient statistics for analysis of the lateral diffusion of lipids, cholesterol molecules, and peptides. Figure 4.5 (panels A, B, C) shows the mean squared displacement (MSD) for each membrane component for each leaflet from the unwrapped trajectories vs time. Ballistic motion can be seen at short times and diffusive motion can be seen starting between 100 ns and 1 μs. The diffusion coefficients were calculated using Einstein's relation and are shown in Table 4.1. Since WT1, WT2, and R694L had the same bilayer composition and include a 27-residue peptide, theoretically the diffusion measurements should be the same. However, as seen in Table 4.1, D_{lat} (10^{-9} cm^2/s) varies from 0.41 to 7.49 for DPPC, 0.47 to 8.90 for cholesterol, and 0.20 to 1.83 for the protein. WT1 and WT2 differed only in their initial coordinates, yet WT1 had 3x faster diffusion in UL for both bilayer components and 5x faster diffusion in LL for both, compared to WT2. Also, all three systems had higher diffusion coefficients in the upper leaflets, compared to the lower leaflets.

The lateral diffusion coefficients measured here are consistent with experimental measurements on membranes with similar compositions, although experiments can be highly variable. Experimental measurements of lateral diffusion on model bilayers are sensitive to temperature, lipid composition and phase, cholesterol and protein content, ion concentration, hydration level, and timescale of measurement. Quasielastic neutron scattering covers picosecond timescales of diffusion, while fluorescence correlation spectroscopy (FCS) and fluorescence recovery after photobleaching (FRAP) cover longer timescales, for instance. NMR measurements on membranes with the same composition used here (50% DPPC and 50% cholesterol) and at 309 K determined the lateral diffusion coefficient to be 33 and 37 x 10^{-9} cm^2/s for DPPC and

cholesterol, respectively [118]. This is within an order of magnitude of the diffusion coefficients here, although the timescales of NMR and simulations are not exactly comparable. Our simulations also support the trend that cholesterol has a higher diffusion rate that DPPC. Another NMR study on bilayers composed of 58% DPPC and 42% cholesterol by Filippov et al. measured lateral diffusion coefficients of 25 to 75 x 10^{-9} cm^2/s for DPPC at 308 K and 313 K, respectively [119]. Scherfeld et al. measured 4.5 x 10^{-9} cm^2/s for a dye using FCS at room temperature on a 50/50 DPPC/chol model bilayer [120]. Computational studies on diffusion of model bilayers are also sensitive to the same parameters as above, as well as force field, system size, area per lipid, MD parameters, and simulation length. Previous studies that measured lateral diffusion coefficients of lipids, cholesterol molecules, and/or proteins in model bilayers were performed for up to 150 ns in cholesterol-free bilayers and up to 500 ns in cholesterol-containing bilayers with all-atom MD [108, 121–125]. Falck et al. simulated a membrane with the same composition at 323 K and measured D_{lat} as 2 x 10^{-9} cm^2/s for both DPPC and cholesterol. However, the authors cautioned that the 100 ns simulation was probably not sufficiently long for diffusion measurements on a system with 50% cholesterol [109]. Although it is complicated to directly compare experimental and computational D_{lat} measurements of similar systems, our coefficients are comparable to both.

Also in Figure 4.5 (panels D, E, F) are plots of mean squared angular displacement (MSAD) for each membrane component for each leaflet vs time. The rotational diffusivities had the same trends as D_{lat}: WT1 had faster rotational motion that WT2 or R694L and the upper leaflets were faster than the lower leaflets. Both the MSD and MSAD for the peptides (C and F) show a sudden drop in diffusivity from 10^3 to 10^4 ns; this is due to the lack of statistics for the peptide at this long timescale and not to a dramatic change in diffusivity. There are 74 to 91 cholesterol or DPPC

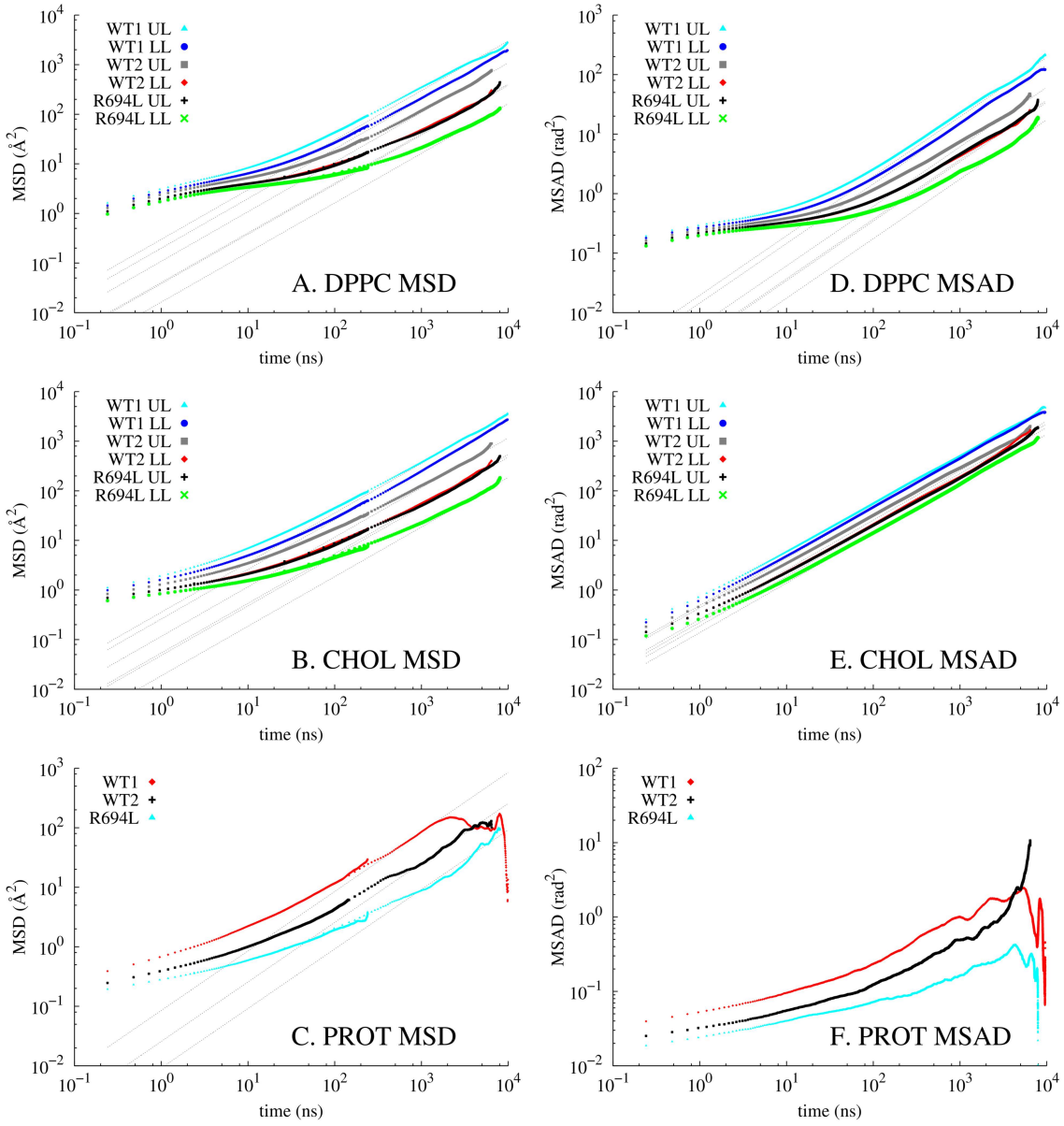

Figure 4.5: Mean-squared displacements vs time for A) DPPC, B) cholesterol (excluding flipflops), and C) the WT or mutant MSD. Dashed lines show best fits of $4D_T t$ for 1000< t <5000 ns (100< t <2500 ns for peptide) to extract the translational diffusion constant. Orientational mean-squared displacements vs time for D) DPPC, E) cholesterol (excluding flipflops), and F) the WT or mutant MSD. Dashed lines show best fits of $2D_R t$ for 1000< t <5000 ns (100< t <2500 ns for peptide) to extract the rotational diffusion constant. In A, B, D, E, "UL" and "LL" refer to upper and lower leaflets, respectively. C and F both show a sudden drop in diffusivity from 10^3 to 10^4 due to the lack of statistics for the peptide at long timescales.

Table 4.1: Species diffusion coefficients

Diffusion coefficients		**Translational (10^{-9} cm^2/s)**			**Rotational (10^{-2} rad^2/ns)**		
System		**WT1**	**WT2**	**R694L**	**WT1**	**WT2**	**R694L**
Protein		1.83	0.63	0.20	0.040	0.019	0.0063
DPPC	UL	7.49	2.71	0.95	0.98	0.31	0.19
	LL	4.98	1.01	0.41	0.73	0.17	0.087
Cholesterol	UL	8.90	2.90	1.20	24.4	12.8	1.44
	LL	6.50	1.34	0.47	22.0	11.1	6.91
Temperature (K)		306.7 ±1.22	302.6 ±1.22	302.2 ±1.24	306.7 ±1.22	302.6 ±1.22	302.2 ±1.24

molecules in a leaflet, so the other graphs do not exhibit such a drop in diffusivity.

4.3.4 Inter-system discrepancies in diffusion manifested by interdependent variables

The inter-system trends of lateral and rotational diffusivities described above correlated to the observed temperature in each system. The cooler systems, WT2 and R694L, had lower diffusivities than the warmer system, WT1. As mentioned in Methods, observed system temperatures (306.7, 302.6, and 302.2 K for WT1, WT2, and R694L respectively) were lower than the setpoint (310 K), as seen in Appendix Figure 4.9. This was a result of the specific choice of MD parameters (no removal of center-of-mass motion and Nosé-Hoover thermostat) to observe true displacements and to avoid fictitious forces. Besides diffusion, the temperature differences between systems resulted in changes in unit cell area, lipid order parameters, and membrane thickness, which we now examine.

The unit cell area, defined by MD periodic boundary conditions, was used to determine membrane area. The systems had two lipid components and a peptide, so

determining area per lipid is neither exact nor trivial. Figure 4.6 shows the unit cell area over the simulation time for all 3 systems. The WT1 system had a stable unit cell area of ∼69 nm^2 for the entire trajectory, which correlated with its stable system temperature. The WT2 and R694L systems relaxed into smaller unit cell areas of ∼64 nm^2, after ∼3 μs, as a result of system cooling. Also evident in Figure 4.6 is when the R694L system occasionally experienced undulatory motions for up to 1 μs in duration, also shown in a snapshot in Figure 4.6. These undulatory motions are represented in the graph by extreme changes in unit cell area that are not reflective of the "true" membrane surface area at that time and have been seen in other studies as well [126]. Mean squared displacements measured during undulatory and nonundulatory periods for the R694L trajectory revealed no correlation with diffusion. Therefore, the lowest diffusion coefficients of R694L cannot be attributed to undulations along with temperature.

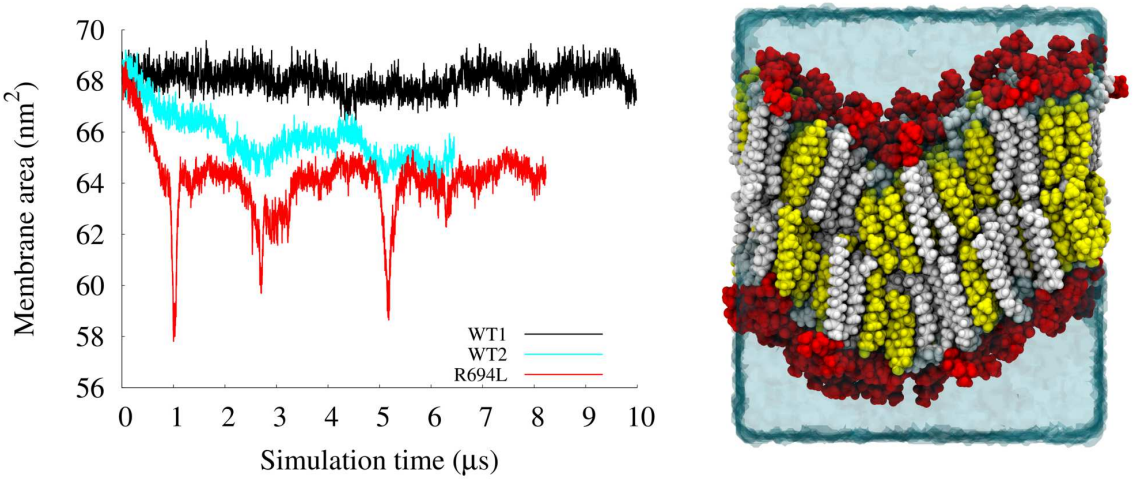

Figure 4.6: Left, total unit cell area during μs simulations. Right, snapshot of R694L with undulation motion.

Lipid tail order parameters in all 3 systems over the entire trajectories also reflect

system cooling and resulting area compression, but are overall similar to other, shorter studies of similar bilayers [107]. As shown in Figure 4.7, both leaflets of WT1 had lower S_{CD} than the leaflets of WT2 or R694L. This indicates that both leaflets of WT1 were less ordered (or more disordered) than the other systems. Also, the WT1 upper leaflet was more disordered (lower S_{CD}) than its lower leaflet. Both WT2 and R694L also had more disordered upper leaflets than their lower leaflets. These trends correlate with the trends in D_{lat}: R694L had the highest S_{CD} and the slowest diffusion while WT1 had the lowest S_{CD} and the fastest diffusion.

The lipid tail order parameters also showed changes with time. Appendix Figure 4.14 shows S_{CD} of each leaflet of R694L over every 2 μs. The upper leaflet increased in order from the first 2 μs compared to the rest of the trajectory, which correlated to initial equilibration of R694L to a new configuration at a lower temperature and a lower total unit cell area. The same trends in S_{CD} occurred for WT2 in both leaflets, as seen in Appendix Figure 4.15, which also required a few μs to relax into a smaller unit cell area. WT1 increased in order from 4 to 6 μs but then decreased to its initial order, as seen in Appendix Figure 4.16. As expected, unit cell area and DPPC order parameters are also correlated with membrane thickness. Histograms of average membrane thickness in 4 x 4 $Å^2$ patches for each system is shown in Appendix Figure 4.17. The more ordered systems, WT2 and R694L, had the thickest membranes and the most disordered system, WT1, had the thinnest membrane.

The slight cooling of the simulations, resulting in $\sim$4 K difference between WT1 and WT2/R694L systems, resulted in large differences in membrane area, diffusion, membrane thickness, and lipid order. The cooler systems, WT2 and R694L, had 5% smaller area and $\sim$10x lower diffusion. The correlation between temperature and membrane area, diffusion, membrane thickness, and lipid order have been observed in other systems [127]. The simulations here support these observations, although

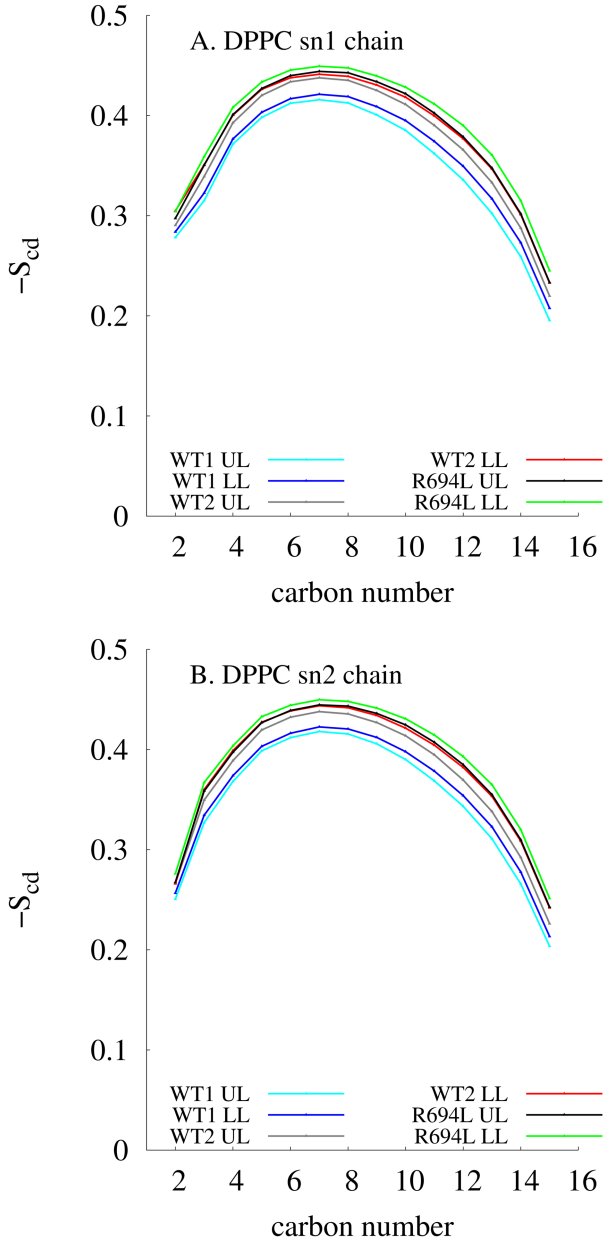

Figure 4.7: DPPC order parameters for each chain over entire trajectory. "UL" and "LL" refer to upper and lower leaflets, respectively.

the temperature differences were the unintentional result of specific MD parameters.

4.3.5 Discrepancies in inter-leaflet diffusion examined

All three systems had faster diffusion in their upper leaflets compared to their lower leaflets. Both WT1 and WT2 systems had a stable water defect in the lower leaflets, as discussed earlier (Appendix Figure 4.12) and in Aim I. Since the lower leaflets of all 3 systems had smaller diffusion coefficients than the upper leaflets, the trend cannot be attributed to the water defect. However, all 3 peptides had the same sequence from residues 15 to 27, including two terminal arginines, which may influence cholesterol or lipid movement. We did not specifically test for this, though.

4.4 Cholesterol flip-flops are rare events

On the μs timescale, several cholesterol molecules attempted to "flipflop" from one leaflet to the other in WT1 and R694L. (No DPPC molecule underwent this motion in any of the systems.) Figure 4.8 shows the z-position and tilt angle of the 3 flipfloping cholesterol molecules vs simulation time. The WT1 system had two such molecules: one flipped quickly to a different leaflet in $\sim$30 ns (A) but the other lingered in the membrane interface for over 2 μs before returning to its original leaflet (B). The quickly flipflopping cholesterol tilted 0.5 ns before it changed z-position; the tilt continued to increase until it reached a maximum. The cholesterol reached the other leaflet 15 ns after reaching the maximum tilt. The cholesterol in panel B remained in the membrane interface for over 2 μs, and then tilted 75 ns before returning to the original leaflet. Jo et al. calculated the PMF of cholesterol flipflop in DPPC and stated that a cholesterol tilts before a change in position [128]. The quickly flipflopping cholesterol of WT1 supported this, but the other cholesterol in WT1 (B) did not tilt first before moving. R694L had one cholesterol (C) that remained in the

membrane interface for the majority of the simulation. This may be because it did not achieve a high enough tilt (like the cholesterol in A and D). Although we have not thoroughly investigated this process, these 3 rare events suggest that a cholesterol will remain in the membrane interface unless it is able to increase its tilt by over 150° and immediately flipflop to the other leaflet. The 3 flipflops here, on a ns to μs timescale in a liquid ordered membrane, do not agree with Choubey et al. [129], who observed faster and more numerous events in a membrane suggested to be at an induced phase transition. However, our simulations may have had a higher free energy barrier to flipflop due to the higher percentage of cholesterol. Compared to experiments, the timescales for cholesterol flipflops here and in Choubey et al. are still considered fast, although there are not necessarily overlaps in timescales of measurements. A recent update to the cholesterol force field parameters in CHARMM36 improved accuracy regarding parallel orientations of cholesterol in the membrane interface; simulations on Anton, such as these here, were not able to utilize them [79]. It is unknown how they would affect the cholesterol flipflopping process.

The flipflopping cholesterols were removed from the diffusion measurements in Table 4.1 because the flipflopping motion complicates the meaning of a lateral diffusion measurement. Including or excluding these 3 cholesterols did not significantly change D_{lat}. WT1 had the same coefficient for cholesterol in the upper leaflet and a decrease from 6.53 to 6.50 x 10^{-9} cm^2/s in the lower leaflet with cholesterol flipflop exclusion. Since R694L had one cholesterol that remained in the membrane interface for the majority of the simulation, the diffusion coefficient in the lower leaflet decreased from 1.02 to 0.47 x 10^{-9} cm^2/s with the quickly-moving cholesterol removed from measurement. It remains unknown if, and to what extent, the cholesterols that remained in the membrane interface in WT1 and R694L influenced DPPC or leaflet motion.

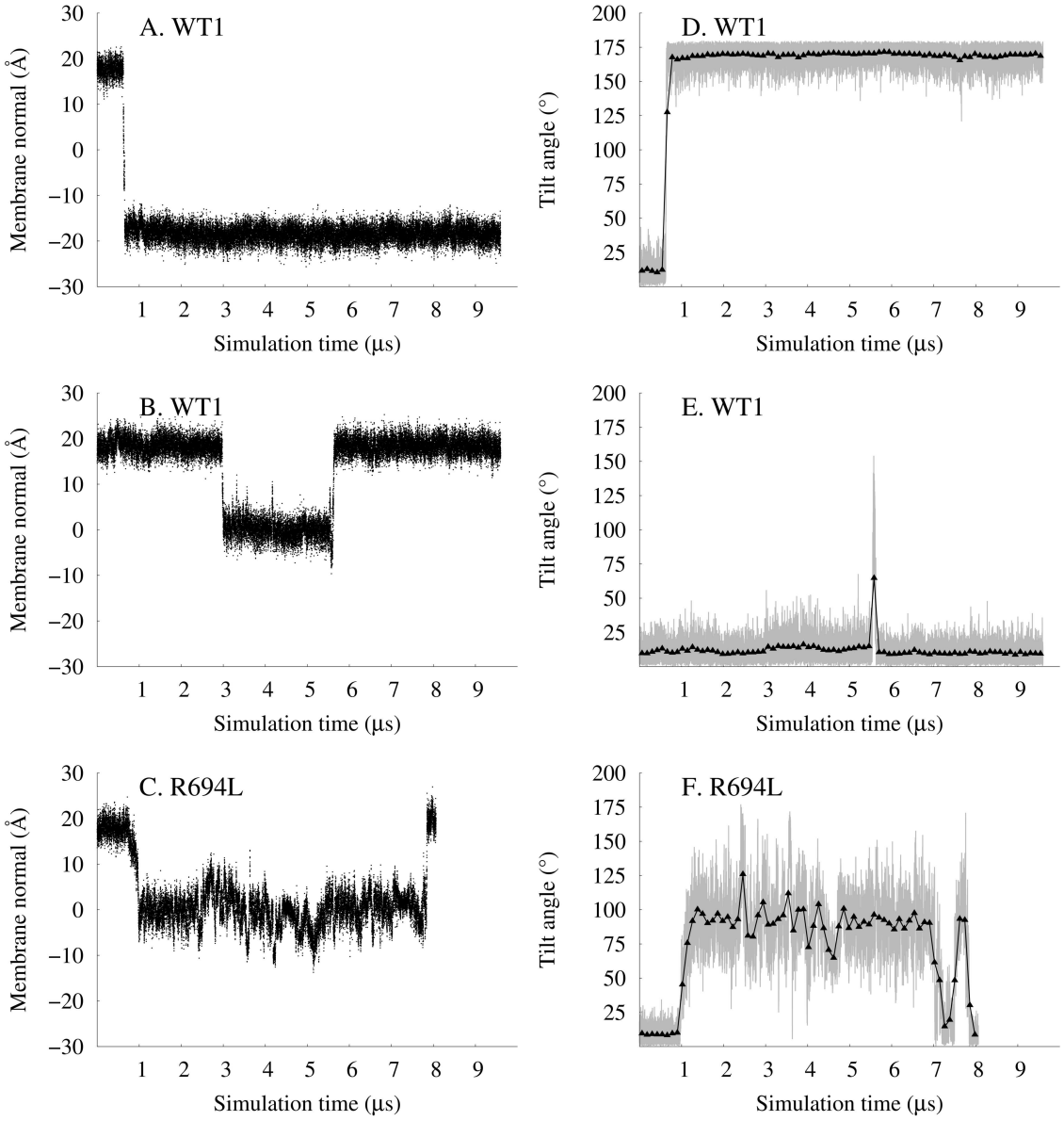

Figure 4.8: Z-positions of oxygen atoms vs simulation time in μs of A, B) 2 cholesterol molecules from WT1 and C) one cholesterol molecule from R694L that flip-flop during the trajectories. D, E, F) Tilt of cholesterol ring with respect to membrane normal vs simulation time, corresponding to cholesterol molecules in (A, B, and C).

4.5 Summary and conclusions of Aim II

Simulations of WT1, WT2, and R694L for 9.98, 6.45 and 8.06 μs MD allowed the model viral bilayers of ∼50% DPPC and ∼50% cholesterol to better configurational sample their phase space. This timescale allowed calculation of diffusion coefficients, which agree with experiments, however inter-system differences are a result of temperature differences of 0.4 to 4.5 K. Temperature control for simulations on the 10 μs timescale therefore requires careful attention as these simulations show for the first time, to the best of our knowledge, that small temperature differences result in large differences in membrane area, lateral and rotational diffusion, lipid order, and membrane thickness at temperatures between 300 and 310 K. The simulations did not achieve complete ergodicity, which suggests that bilayers with 50% cholesterol require longer than 10 μs for proper sampling. A recent umbrella sampling calculation of the binding of a small peptide to a bilayer came to the same conclusion when it required windows of 1.5 μs [130].

4.6 Appendix

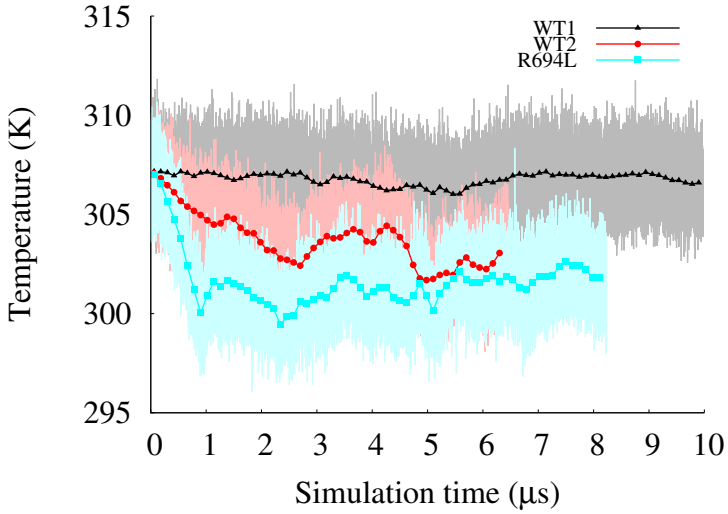

Figure 4.9: Simulation temperature (K) over time for WT1, WT2, and R694L.

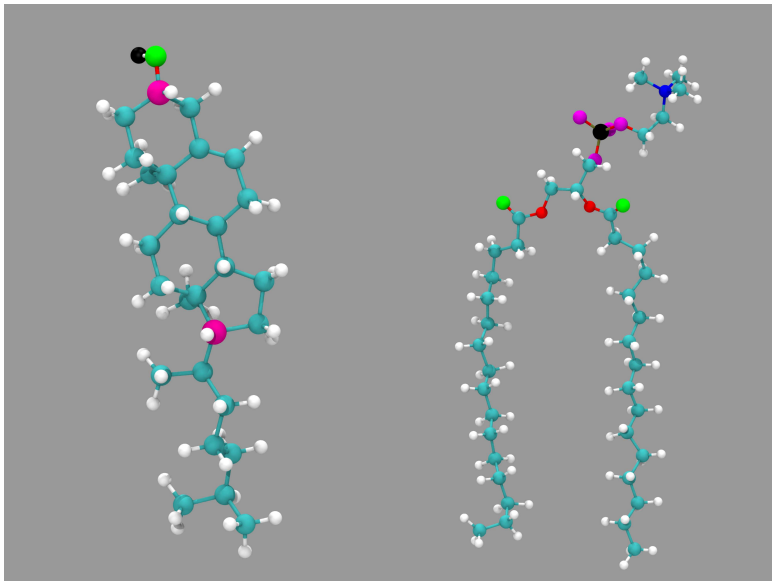

Figure 4.10: All-atom renderings of a cholesterol (left) and a DPPC (right) molecule to highlight atoms selected for analysis of cholesterol tilt, 2D maps of average mass, and radial distribution function. The atoms not used for analysis are highlighted by element: hydrogen in white, carbon in light blue, oxygen in red, and nitrogen in dark blue. Atoms used in analysis are in different colors: in cholesterol, C3 and C17 in pink, hydroxyl oxygen O3 in green, and hydroxyl hydrogen H3′ in black; in DPPC, phosphorous atom in black, phosphate oxygens in pink, and carbonyl oxygens in green.

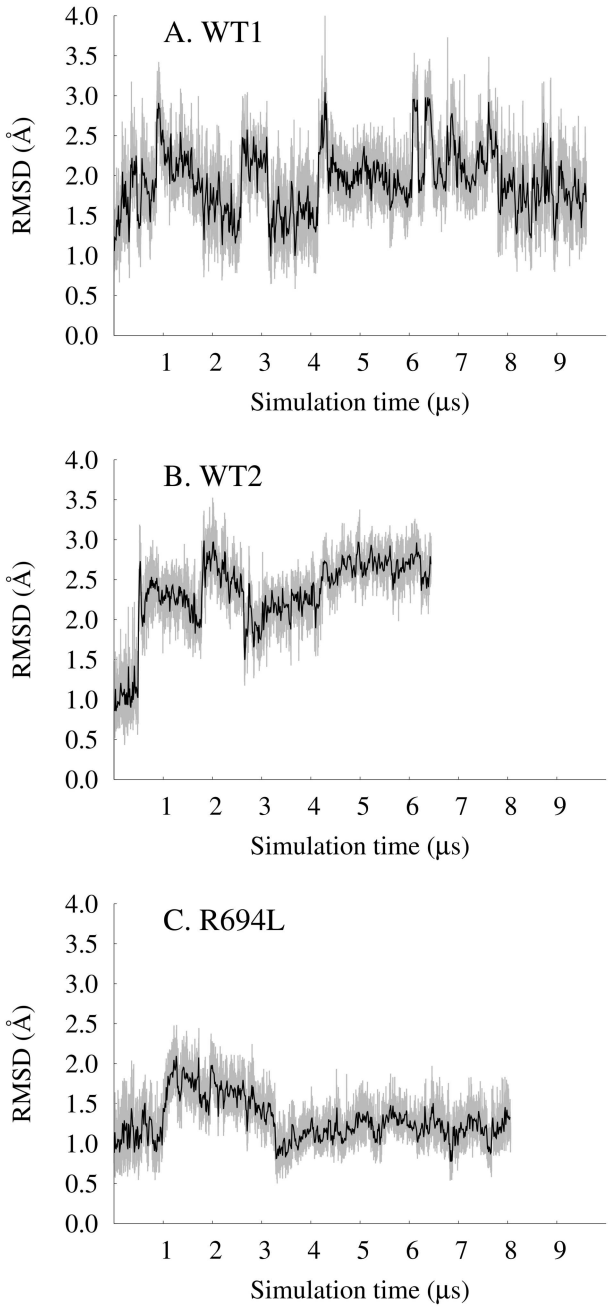

Figure 4.11: RMSD of backbone compared to t = 0 in Å vs simulation time in μs.

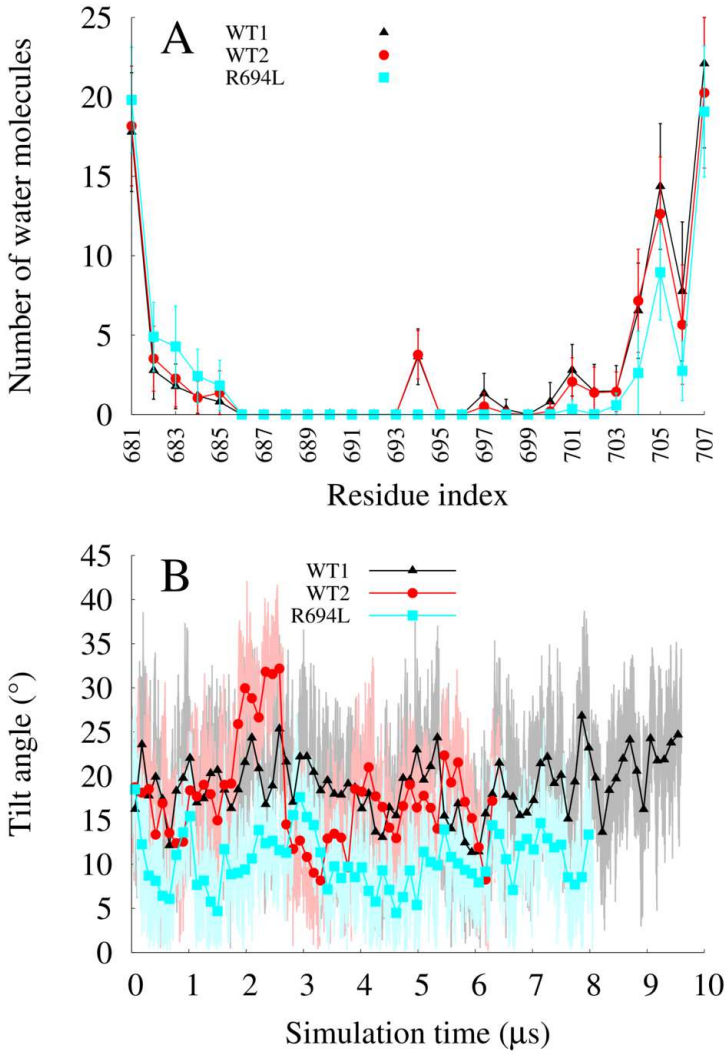

Figure 4.12: Top, number of water molecules within 4 Å of protein vs amino acid that each water molecule is uniquely attributed to during MD. All statistics are from the entire trajectories and error bars represent standard deviation. Bottom, tilt angle of helix in degrees vs simulation time in μs.

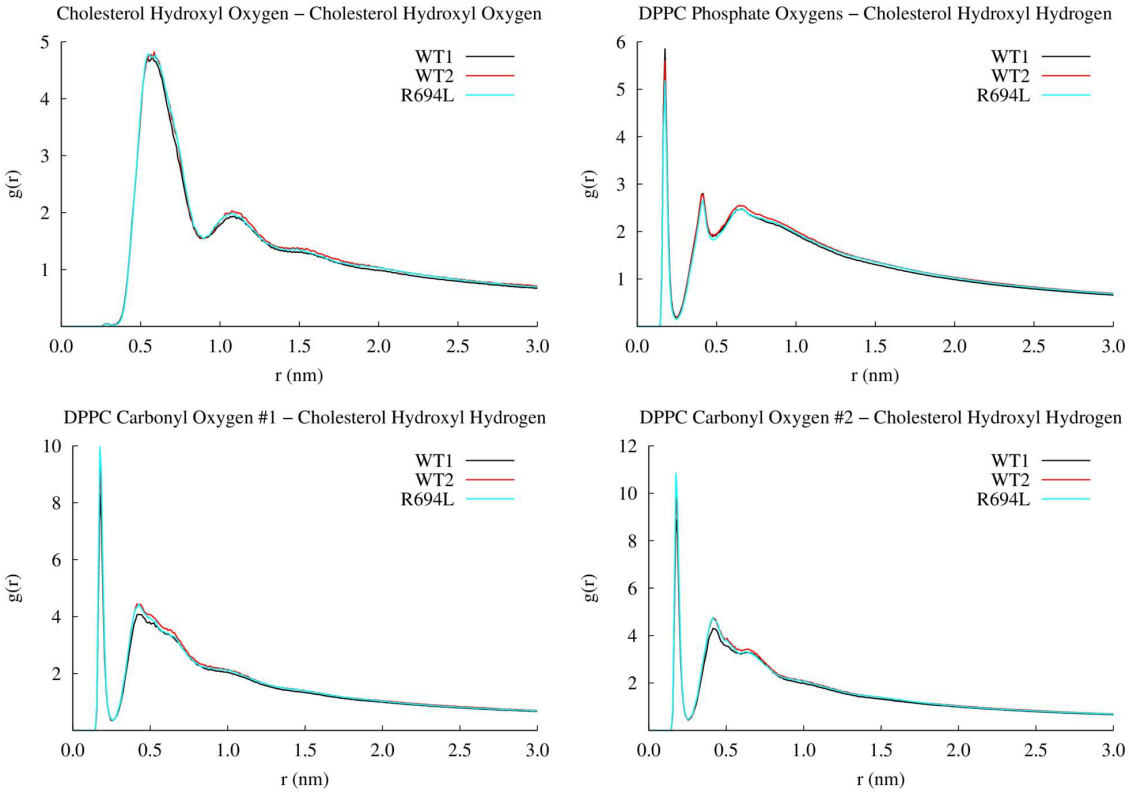

Figure 4.13: Radial distribution systems for cholesterol hydroxyl oxygen with respect to itself and cholesterol hydroxyl hydrogen with respect to either the lipid phosphate oxygen atoms or the lipid carbonyl oxygen atoms for WT1, WT2, and R694L.

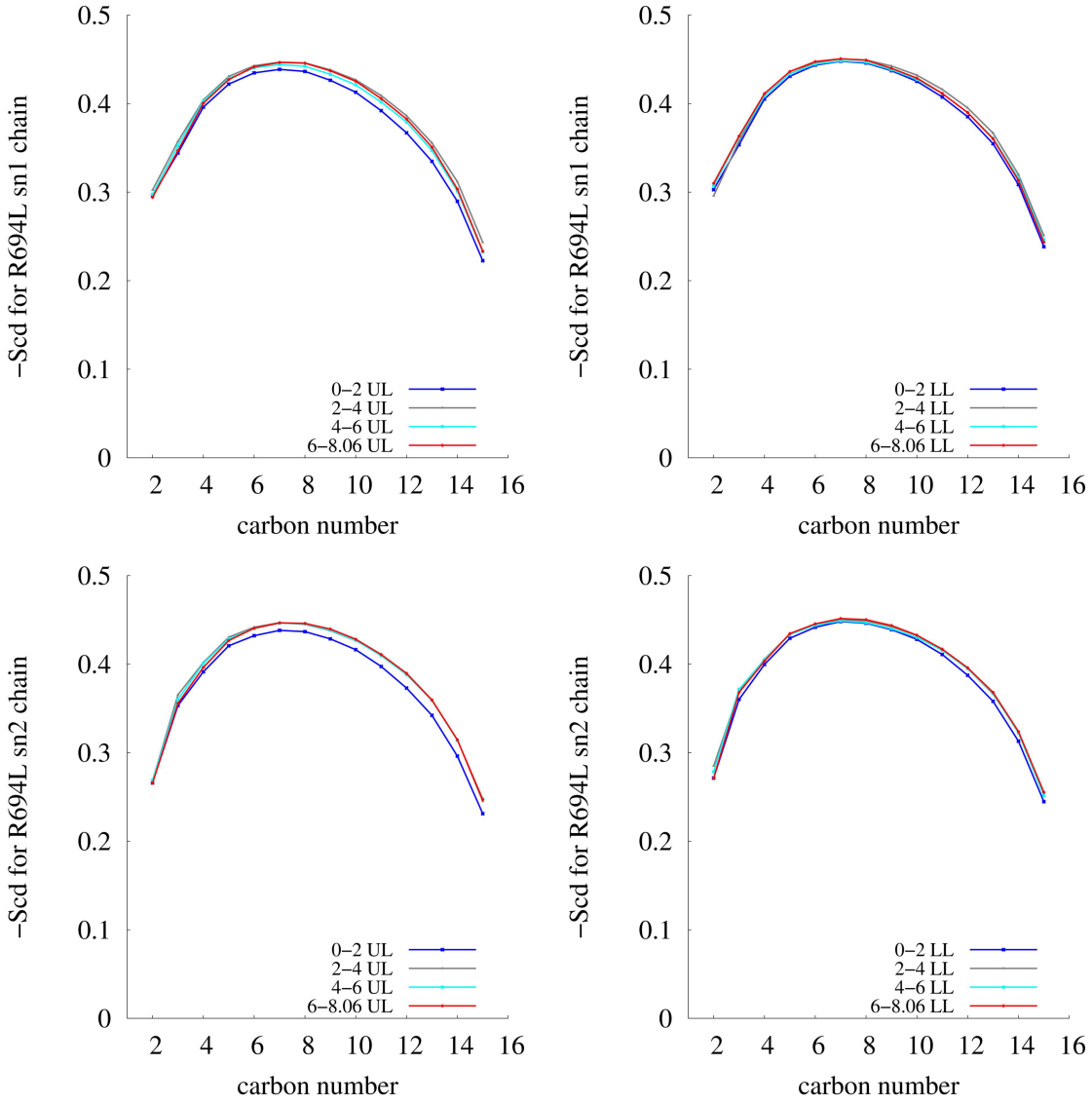

Figure 4.14: DPPC tail order parameters (sn1 and sn2) vs carbon atom index every 2 μs for upper and lower leaflets of R694L.

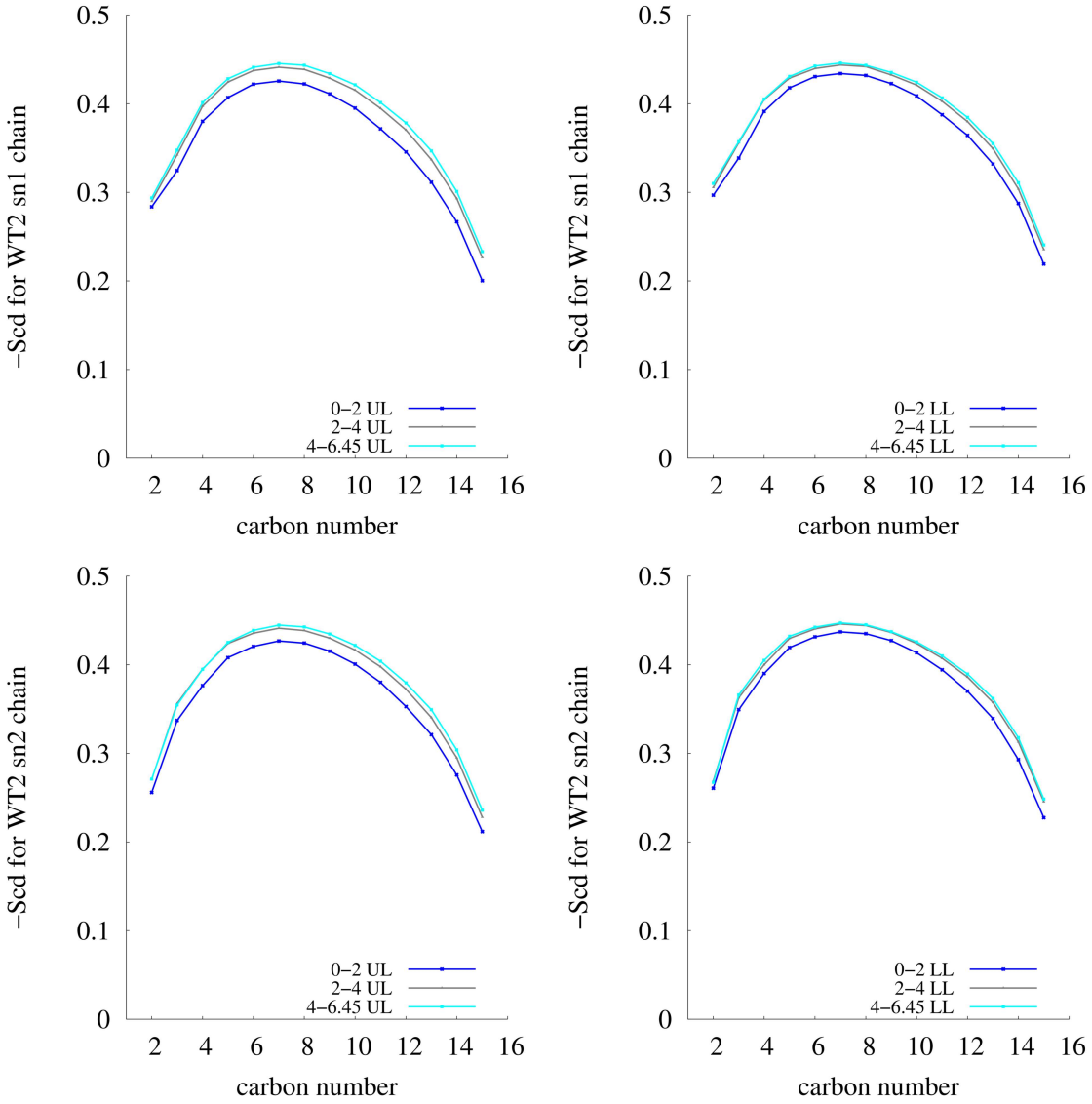

Figure 4.15: DPPC tail order parameters (sn1 and sn2) vs carbon atom index every 2 μs for upper and lower leaflets of WT2.

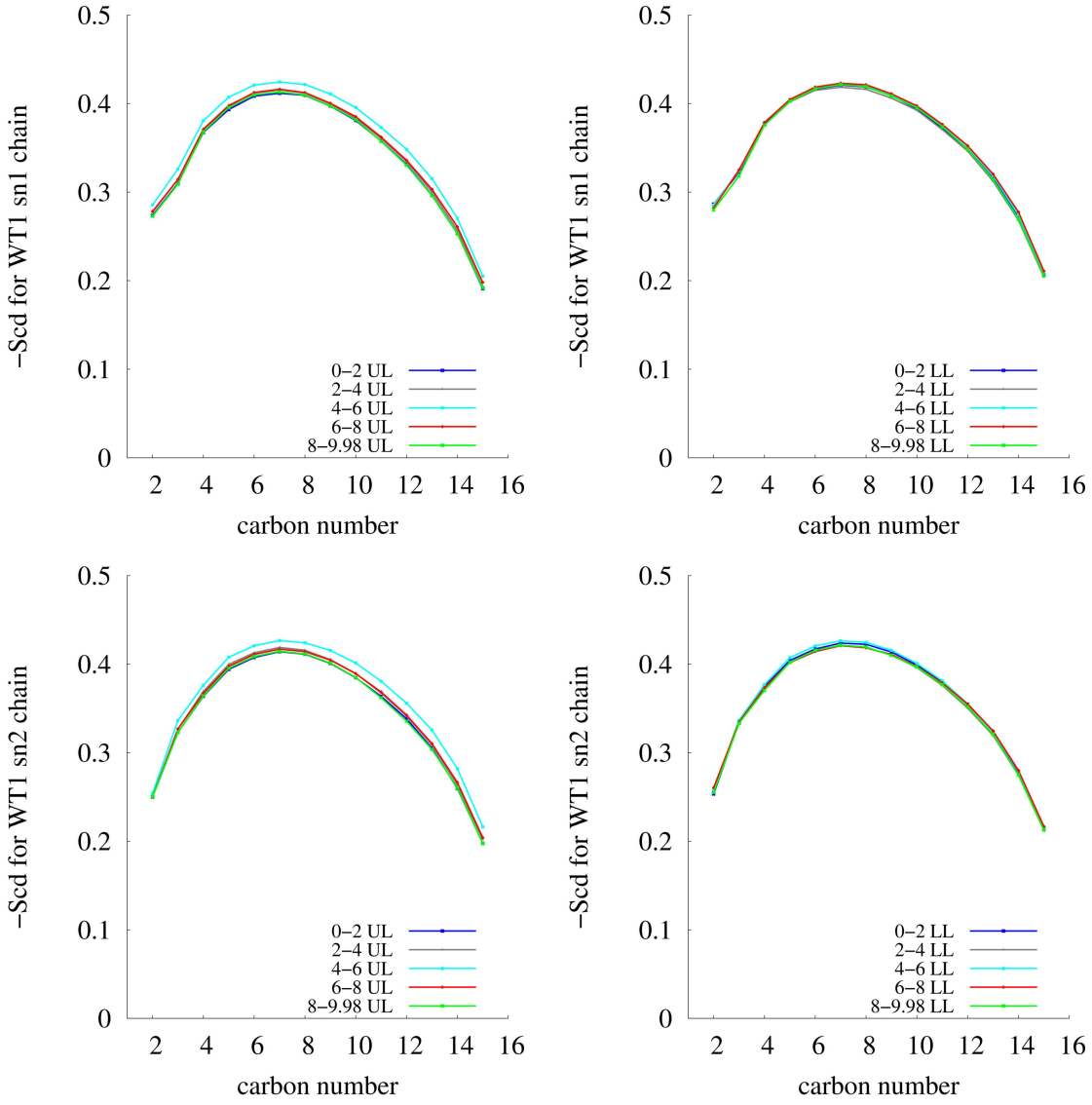

Figure 4.16: DPPC tail order parameters (sn1 and sn2) vs carbon atom index every 2 μs for upper and lower leaflets of WT1.

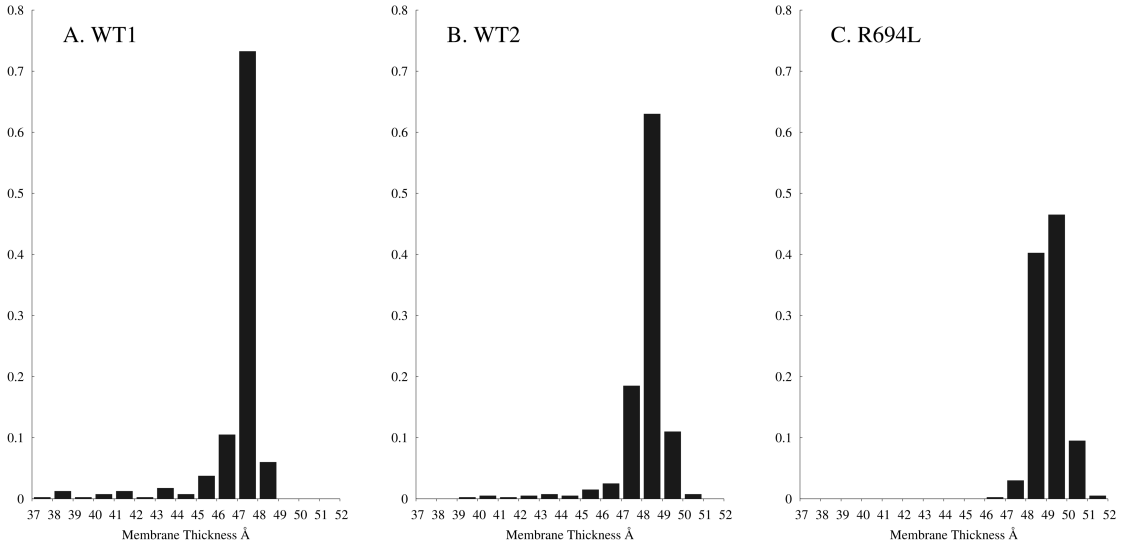

Figure 4.17: Histograms of average membrane thickness in 4 x 4 Å^2 patches for WT1, WT2, and R694L systems.

5. Aim III: Modeling of HIV-1 gp41 MPER-MSD trimer in a cholesterol-containing bilayer

5.1 Introduction to Aim III

Since the fusion-active spikes of HIV-1 are trimeric (trimers of dimers gp120 and gp41) and gp41 MSD contains the common α-helical transmembrane interaction motif GXXXG, it is reasonable to suppose that MSD exists as a trimer either in the native, unliganded spike or at some point during fusion. As mentioned, there are currently no crystal or NMR structures of the MSD. In this thesis, we have used comprehensive simulations with a segment-based approach to model the monomeric MSD peptide in a viral bilayer (Aim I) and to investigate the ergodicity of the model viral bilayer on μs timescales (Aim II). Information obtained from Aims I and II enabled us to now rationally construct a model of a MSD trimer. Model building of membrane proteins is sometimes performed in hydrophobic slabs, coarse-grained membranes, or all-atom membranes; however, the enumeration of the MSD was performed in vacuum due to the specific attributes of the MSD-membrane environment. Enumeration in a hydrophobic slab or implicit membrane would neglect the importance of the water defect and the requirement of cholesterol for fusion [131], as studied in Aim I. Enumeration of all trimer models in an all-atom membrane is too computationally expensive as the slow diffusion of lipids and cholesterol, studied in Aim II, requires at least μs long trajectories for partial ergodicity of the model membrane. Enumeration in vacuum is therefore the only acceptable approach for the HIV-1 gp41 MSD. Also, van der Waals forces have been found to be more important for transmembrane helical interactions than ionic or hydrophobic forces [132–135], further supporting a vacuum enumeration. Therefore, the MSD trimers were enumerated in vacuum and

then the lowest energy model was equilibrated in a model bilayer.

Finally, we wished to augment this model with structures of the trimeric MPER to produce a MPER-MSD trimer model in a viral bilayer. Since MPER is amphipathic, several experimental structures of MPER exist. Experimental structures of MPER include a kinked α-helix in micelles and a postfusion crystal structure of helical MPER as an extension of 6HB, for instance [24, 26]. In the native, unliganded spike, MPER's epitopes are not accessible to antibodies. Recent cryo-EM structures of the unliganded spike show a "stalk" form for the exposed ectodomain of gp41 [19, 136]. Although there is no consensus about how MPER plays a role in fusion, the putative prefusion structure has been proposed to be an α-helical trimer parallel to membrane normal with the epitope regions located in the trimer interface. A similar structure was recently crystallized with an isoleucine zipper motif [25]. Also, a recent NMR structure of the MPER trimer in a DPC micelle claims to be a putative prefusion intermediate state [137]. These trimeric MPER structures enabled us to model a prefusion MPER-MSD trimer in a model viral bilayer. With this model, we aimed to answer the following questions:

1. Is the lowest energy MSD trimer consistent with what is known about the conserved residues?

2. Are the MSD and MPER-MSD trimer models stable on the 100 ns timescale?

3. Are the midspan arginines still solvated and do the models provide any clues to a poration mechanism?

5.2 Methods

5.2.1 Enumeration of MSD trimer

A large discrete set of trimer configurations spanning the entire relevant ranges of several internal variables were enumerated in vacuum using an α-helical MSD peptide determined to be a stable structure from previous group work [1]. The peptide was placed parallel to the z-axis and triplicated. The three peptides were then rotated around the z-axis (α, 10 to 360° in increments of 10°), around the x-axis (β, -45 to 45° in cosine distributed increments of 5°), and around the z-axis (γ, -45 to 45° in increments of 5°), resulting in unique transformations specified by Euler angles. The peptides were then rotated to be threefold symmetric around the z-axis and translated equilaterally away from the initial trimer center (5 to 25 Å, in increments of 1 Å). An example *tcl* script for the enumeration is located in Appendix B. This enumeration gave ∼270,000 structures; all structures were sorted by potential energy (vdW and electrostatics) and the lowest energy structure was found. In addition to enumeration of the WT peptide with an endoplasmic snorkeling midspan arginine, other enumerations were performed for the MSD with an exoplasmic snorkeling midspan arginine, with the WT BaL sequence, and with neutral N- and C- termini.

5.2.2 Setup and equilibration of MSD trimer

In order to generate a system with the lowest-energy trimer in a bilayer, the WT1 system with monomeric MSD in a bilayer was used as a starting point. Two other monomeric MSD peptides were placed vertically above the bilayer. This system was minimized and equilibrated for 2 ns NPT MD with protein backbones constrained. One of the peptides above the bilayer was rotated around the z-axis using the rotating constraints feature in NAMD, at a rate of 90° per 100 ps for 100 ps MD. Steered MD (SMD) was used to pull the two peptides individually into the bilayer with a force

constant of 7 kcal/mol/Å^2, a velocity of 1 x 10^{-5} Å/timestep in the z-direction, and for a total of 14 ns. During SMD, the secondary structure of the SMD peptide was restrained with extra bonds, the lipid and cholesterol molecules were restrained in the z-direction, and the remaining soluble peptide was fixed. After SMD, excess water and 4 obtrusive lipid and cholesterol molecules were deleted and the system was briefly equilibrated for 2 ns. Next, targeted MD (TMD) was used to pull the three peptides into the lowest-energy vacuum structure with a force constant of 200 kcal/mol/Å^2 for 5 ns. The resultant trimer model had an RMSD of less than 0.5 Å with respect to the vacuum trimer. The trimer-membrane system was run for 100 ns NPT MD with the backbone constrained to allow lipid, cholesterol, and water to relax around the trimer. Then the protein constraints were released and MD was continued for 150 ns. The same parameters as the WT1c system in Chapter 3 were utilized here.

5.2.3 Preliminary MPER simulations

Preliminary simulations of MPER in water were performed using a monomeric NMR structure (PDB 2PV6). The monomeric MPER was solvated in a water box, neutralized, ionized, and minimized. To explore the ensemble of MPER in water, three replicas were equilibrated for 500 ps NPT MD and run for 50 ns NVT MD. The simulations were run at 310 K using CHARMM27, TIP3P water, periodic boundary conditions, and a grid spacing of 2.0 Å. Verlet integration was applied every 2 fs and nonbonded forces were shifted off from 8 to 9 Å.

5.2.4 Setup and equilibration of MPER-MSD trimer

Trimeric MPER was excised from a recent crystal structure (PDB 3G9R), aligned to the N-terminal of trimeric MSD after 100 ns of equilibration described above, cova-

lently bonded, minimized, and equilibrated for 50 ns with constraints on nonhydrogen protein atoms. Then, the constraints were released and the system was run for 133 ns NPT MD. The same parameters as the WT1c system in Chapter 3 were utilized here. This system had 97,825 atoms, including ions, 231 DPPC molecules, 275 cholesterol molecules, 2,520 protein atoms, and ∼15,000 water molecules. The system size was 105 x 100 x 85 $Å^3$.

5.2.5 Production run of MPER-MSD trimer

The MPER-MSD trimer after 133 ns MD was simulated for 10.973 μs NPT MD on the MD-specialized machine, Anton [110, 111]. As before, the CHARMM force field with recent lipid-based corrections and explicit TIP3P water were used [75–78]. Verlet integration was applied every 2 fs and long-range electrostatics were handled with the Gaussian Split Ewald method [112]. The Langevin thermostat was set to 310 K and the Martyna-Tobias-Klein barostat was set to a semi-isotropic pressure of 1 atm [80, 81, 113]. Trajectories were visualized with VMD [74]. See Appendix C for the complete list of configuration parameters.

After ∼11 μs, the system was converted to NAMD format and was equilibrated with the protein backbone constrained. The trimer was then pulled upwards in a constant-velocity SMD simulation for 12 ns where forces were applied to the MPER backbone, the force constant was 10 kcal/mol/$Å^2$, and the velocity was 2 Å/ns. During SMD, lipid molecules farther than 20 Å from the trimer were constrained in the z-direction. NAMD parameters for the SMD simulation were the same as the WT1c system in Chapter 3.

5.2.6 Calculations of observables

The potential energy of the vacuum models was assessed by summation of the Lennard-Jones potential and the Coulombic potential using ϵ, σ, and atomic charges from the CHARMM force field. The radial distribution function was calculated using the *g(r)* plugin in VMD and a δ of 0.1 Å [117]. The pair selections chosen were the cholesterol hydroxyl oxygen and either a particular residue or the K681 NH_3^+ atoms. The lipid order parameters, number of unique waters per residue, maps of membrane thickness, and maps of water penetration were calculated similarly to Chapters 3 and 4.

5.3 Aim III Results

5.3.1 Lowest energy structure of MSD trimer in vacuum is determined from ∼270,000 possible geometric configurations

As a first step towards an equilibrated trimeric MSD, all potential geometric models of the MSD trimer were systematically constructed, resulting in ∼270,000 unique Euler transformations. The α angle varied from 0 to 360°, however the β and γ angles were constrained to be between -45 and 45°. This was based on the assumption that the MSD is a single pass transmembrane helix, with the N-terminus in the upper leaflet, the C-terminus in the lower leaflet, and the improbability of the MSD to be found horizontal in the membrane. It was also assumed that the trimer was three-fold symmetric. The resulting ∼270,000 models were evaluated for potential energy. Figure 5.1 shows the histogram of potential energies for all negative energy structures. The lowest energy structure for the WT sequence with an endoplasmically snorkeling arginine (WTendo in Table 5.1) was found to be a left-handed trimer with a rotation of 310° around z, -45° around x, 40° around z, and each helix translated 7 Å

laterally from the trimer center of mass. Interestingly, the highly conserved GXXXG motifs, postulated to mediate transmembrane helical interactions, were found to be interacting in the vacuum structure, as seen in Figure 5.2.

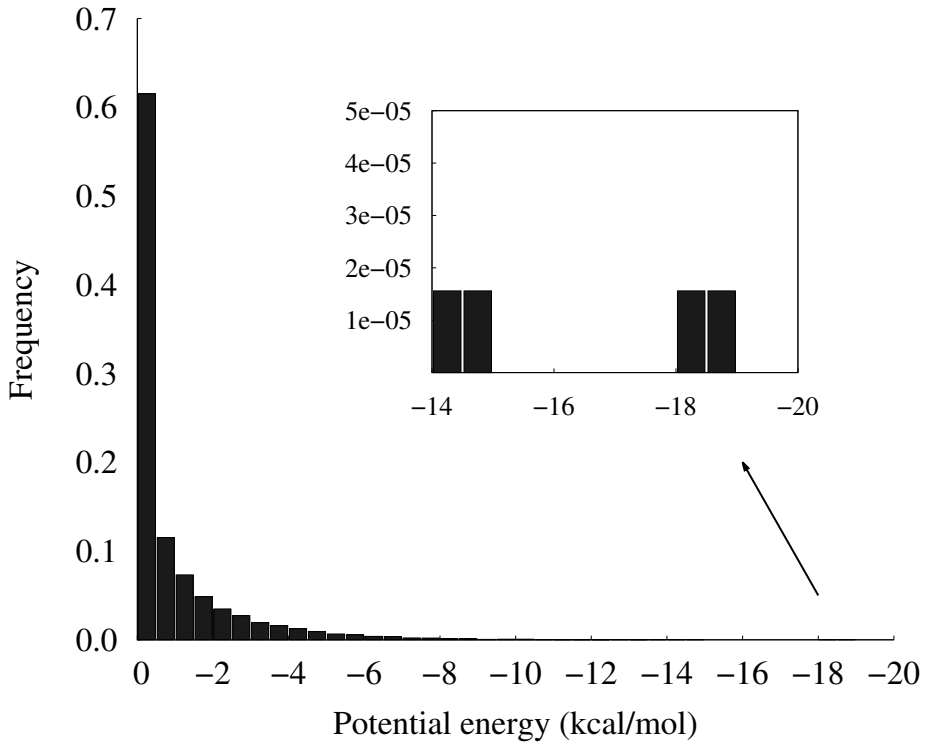

Figure 5.1: Histogram of potential energy for all enumerated WT structures with a endoplasmically snorkeling arginine and potential energy per dimer less than 0 in kcal/mol. The inset shows placement of the 4 lowest energy trimers.

Changing the WT sequence from HxB2 to BaL (difference of 5 residues, see Appendix Table 5.2) did not alter the lowest energy structure appreciably. Neither did changing the terminal atoms from the traditional charged termini (3H-N^+-CH-C=O... for N-terminus and O^--C=O... for C-terminus) to neutral termini (CH3-(C=O)-NH-CH... for N-terminus and CH3-NH-C=O... for C-terminus). (Please refer to NTER, CTER, ACE, and CT3 in the CHARMM force field for further information). The

Table 5.1: Lowest energy Euler transformations for the MSD

Monomer	**Energy per dimer (kcal/mol)**	**Energy per structure (kcal/mol)**	$\alpha(^\circ)$	$\beta(^\circ)$	$\gamma(^\circ)$	**Equilateral distance (Å)**	**Handedness**
WTendo	-18.6	-55.8	310	-45.0	40	7	L
WTBaLendo	-17.8	-53.4	310	-45.0	45	6	L
neutWTendo	-18.3	-54.9	310	-45.0	45	6	L
neutWTBaLendo	-18.1	-54.3	310	-45.0	45	6	L
WTexo	-18.0	-54.0	30	-39.4	-40	7	R
neutWTexo	-17.5	-52.5	170	-39.4	40	8	L

lowest energy trimers with an endoplasmic snorkeling arginine were all left-handed, have the same α and β angles, and similar γ angles. Each monomer was 6 to 7 Å from the trimer center of mass. The WT monomers were also enumerated with an exoplasmic snorkeling arginine, even though previous group work determined that the endoplasmic arginine was lower in energy and more mechanistically realistic with translocon-mediated *env* translation than the exoplasmic snorkeling arginine [1]. The enumeration of the MSD with an exoplasmic snorkeling arginine resulted in a right-handed trimer where R694 reaches towards the upper leaflet. Enumeration of the exoplasmic snorkeling arginine with neutral termini resulted in a left-handed trimer but R694 was not positioned to snorkel and was located horizontally in the middle of the bilayer. This structure is probably not stable in a membrane environment.

5.3.2 MSD trimer is not stable during 150 ns MD

The lowest energy MSD trimer in vacuum with an endoplasmic snorkeling arginine (Figure 5.2) was then placed in a bilayer (50% DPPC and 50% cholesterol) and simulated for 150 ns NPT MD. During the last 50 ns, the trimer in the bilayer continuously increased in RMSD with respect to the vacuum structure, as seen in

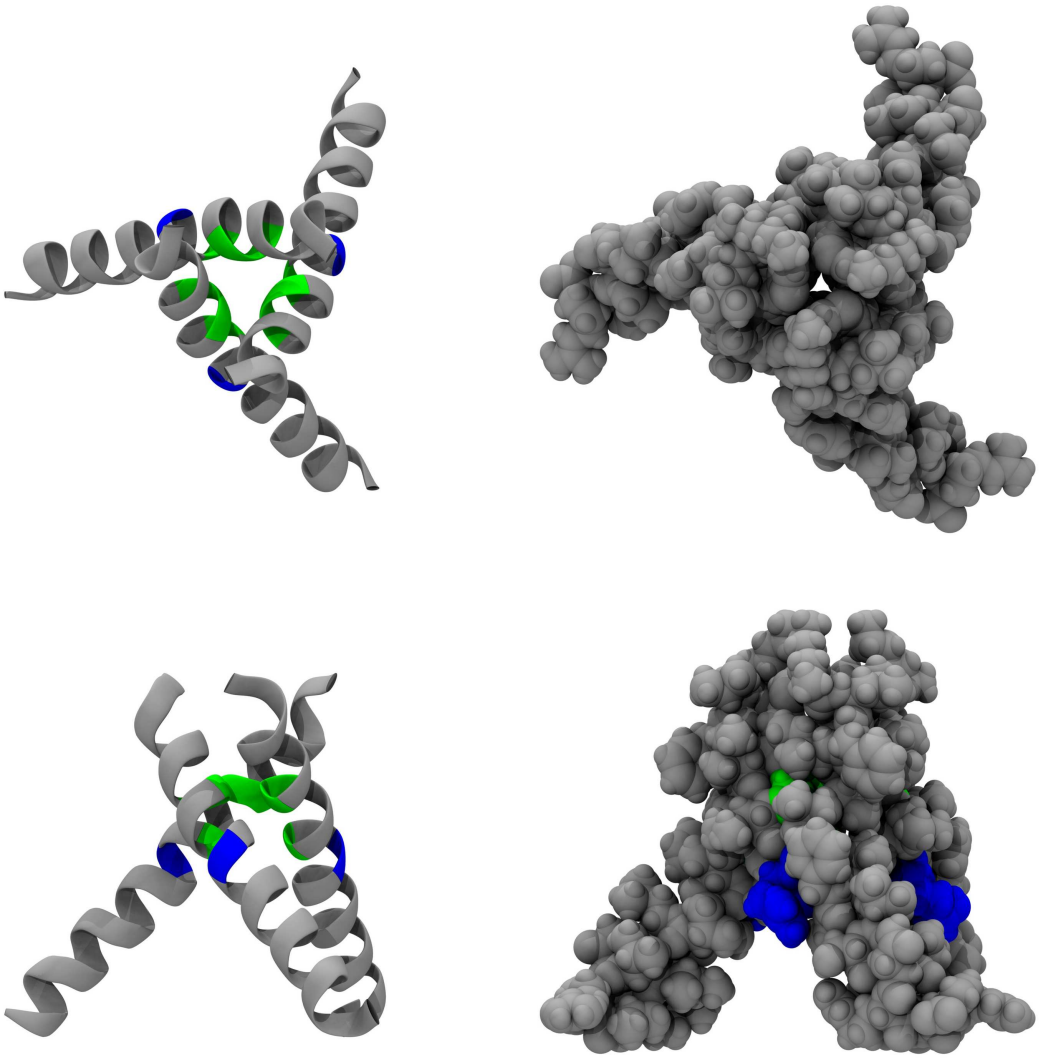

Figure 5.2: Snapshots of the lowest energy MSD trimer model with endoplasmically snorkeling arginines: midspan arginine in blue, glycines in GXXXG motif in green, other residues in silver.

Figure 5.3. This indicated that the trimer is not stable on 150 ns MD timescales. The simulation did not support the GXXXG motif as a trimerization motif or as stronger than lipid/cholesterol-protein interactions. There is also the possibility that the lowest-energy trimer in vacuum is not a low energy structure in membrane nor stable in a membrane environment. The angle of each helix in the vacuum model from

membrane normal was ∼20° greater than the tilt of the monomeric WT peptides in the cholesterol-containing bilayers (Chapter 3). This discrepancy between the ideal tilt of the trimer helices in vacuum and the ideal tilt of the monomeric helix in a bilayer, to accommodate the water defect, was perhaps the reason for the increasing RMSD of the trimer. Also, the MSD trimer may require other segments of gp41 to trimerize, such as MPER, the gp41 segment N-terminal to MSD.

5.3.3 MPER stabilized the MSD trimer for 133 ns MD

Next, it was investigated whether the addition of trimeric MPER, the segment N-terminal to MSD, was necessary to stabilize the MSD trimer in a membrane environment. Therefore, an appropriate trimeric MPER structure was chosen. Preliminary simulations were performed of MPER in a water box to generate a collection of trajectories. Three replicas with different initial velocities were simulated for 50 ns NVT. The three trajectories quickly diverged from the initial NMR structure in ∼5 ns, as seen in Figure 5.4. Each trajectory sampled around a different RMSD with respect to the initial structure. This indicates MPER is quite flexible and probably has highly heterogeneous ensembles in water and membrane. Therefore, MPER cannot be enumerated in water or membrane as straightforward as with the MSD α-helix.

However, there are many experimental structures of soluble or liposome-associated MPER. All structures with only partial MPER (such as merely the CRAC domain) were removed from consideration as well as structures complexed with antibodies. Therefore, the putative prefusion crystal structure of trimeric MPER [25], also left-handed, was covalently added to the N-terminus of the MSD trimer from the branched off simulation of the trimer in bilayer after 100 ns MD with constrained backbones. After equilibration with constraints, the MPER-MSD system was subjected to unconstrained MD for 133 ns, where it stabilized ∼3 Å RMSD, as seen in Figure 5.5. Clearly

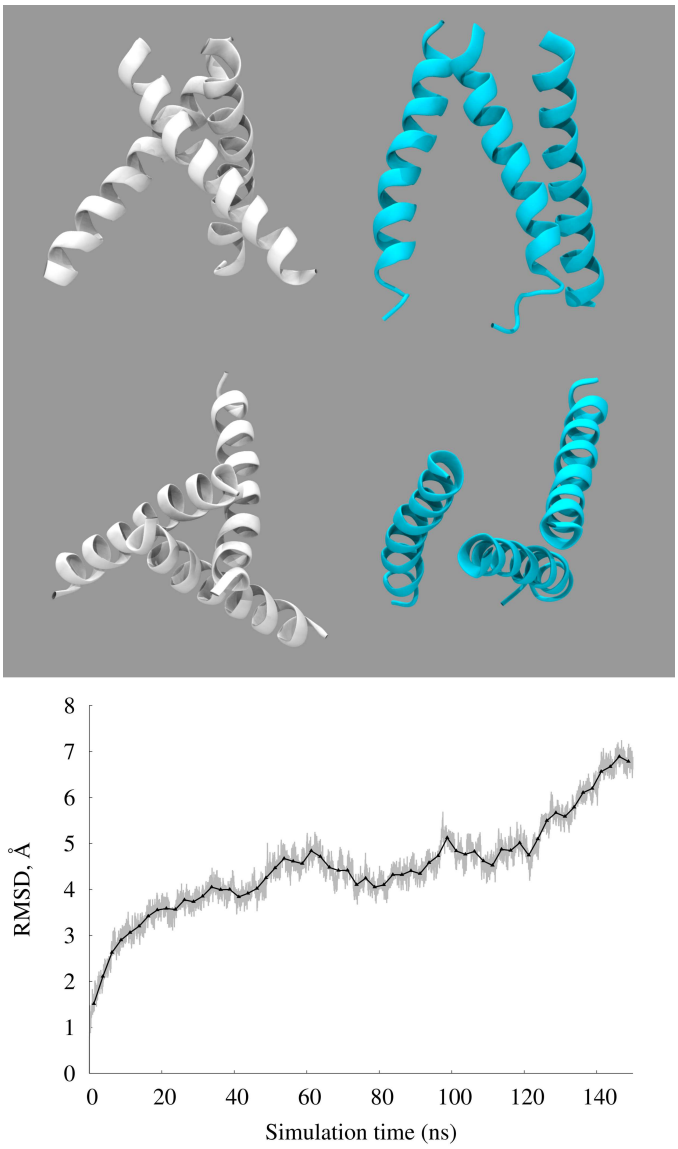

Figure 5.3: In white, vacuum MSD trimer model. In blue, MSD trimer model after 150 ns. Bottom, RMSD vs simulation time.

the addition of trimeric MPER stabilized the trimeric MSD on the 100 ns timescale. Next, the simulation of the unconstrained MPER-MSD trimer was continued for ∼11 μs to determine if the structure was stable on μs timescales.

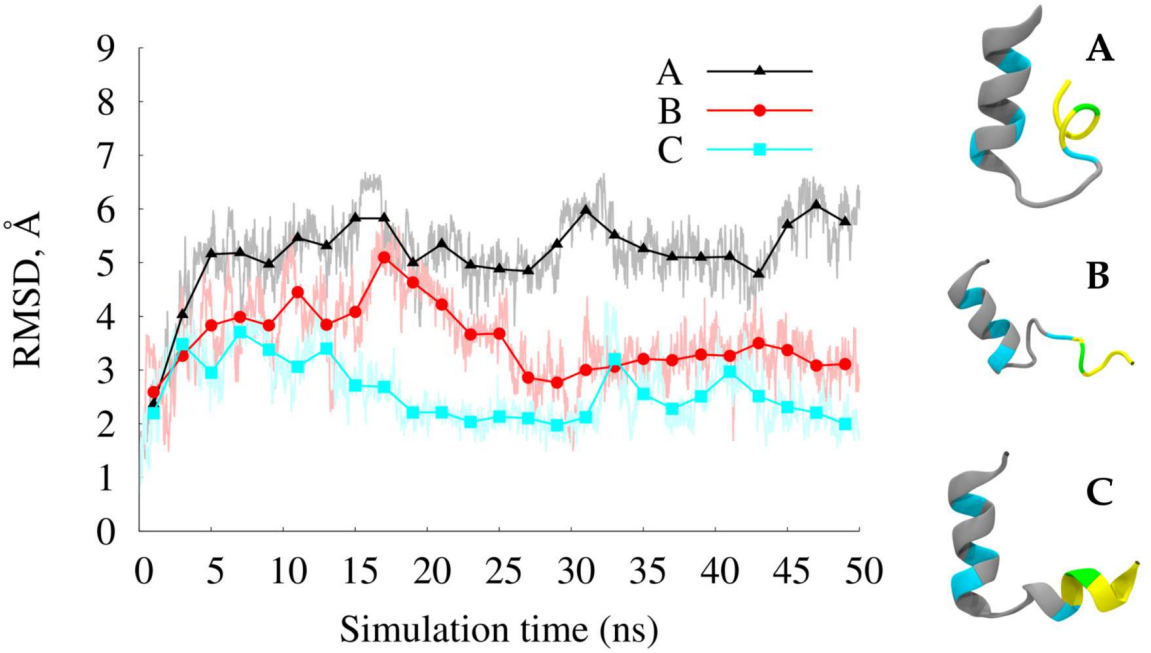

Figure 5.4: RMSD of backbone vs simulation time for three MPER replicas. The configuration after 50 ns is shown on the right with the same color scheme as Figure 1.1: tryptophans in blue (W664, W668, W670, W676), CRAC domain in yellow (L677, W678, Y679, I680, K681), and CRAC tryptophan in green (W678).

5.3.4 MPER-MSD trimer rearranges to new configuration on μs timescale

The MPER-MSD trimer (after 133 ns unconstrained MD) was simulated for an additional ∼11 μs. It took ∼5 μs to settle into a different state, at which point it was stable for the remaining 6 μs, as seen in Figure 5.6. This new configuration was 7 Å RMSD from the initial structure. The main observation is that the GXXXG motifs were no longer interacting, as seen in Figure 5.6, although the N-terminal residues of MPER were still close together. The right-hand panel of Figure 5.6 shows the trimer backbone near K681 (in green), which is the last residue in the CRAC motif and the dividing residue between MPER and MSD. After 11 μs, the three lysines interacted in the trimer interface and cloistered water molecules and a cholesterol molecule. Radial distribution functions ($g(r)$) between the three lysines and cholesterol hydroxyl

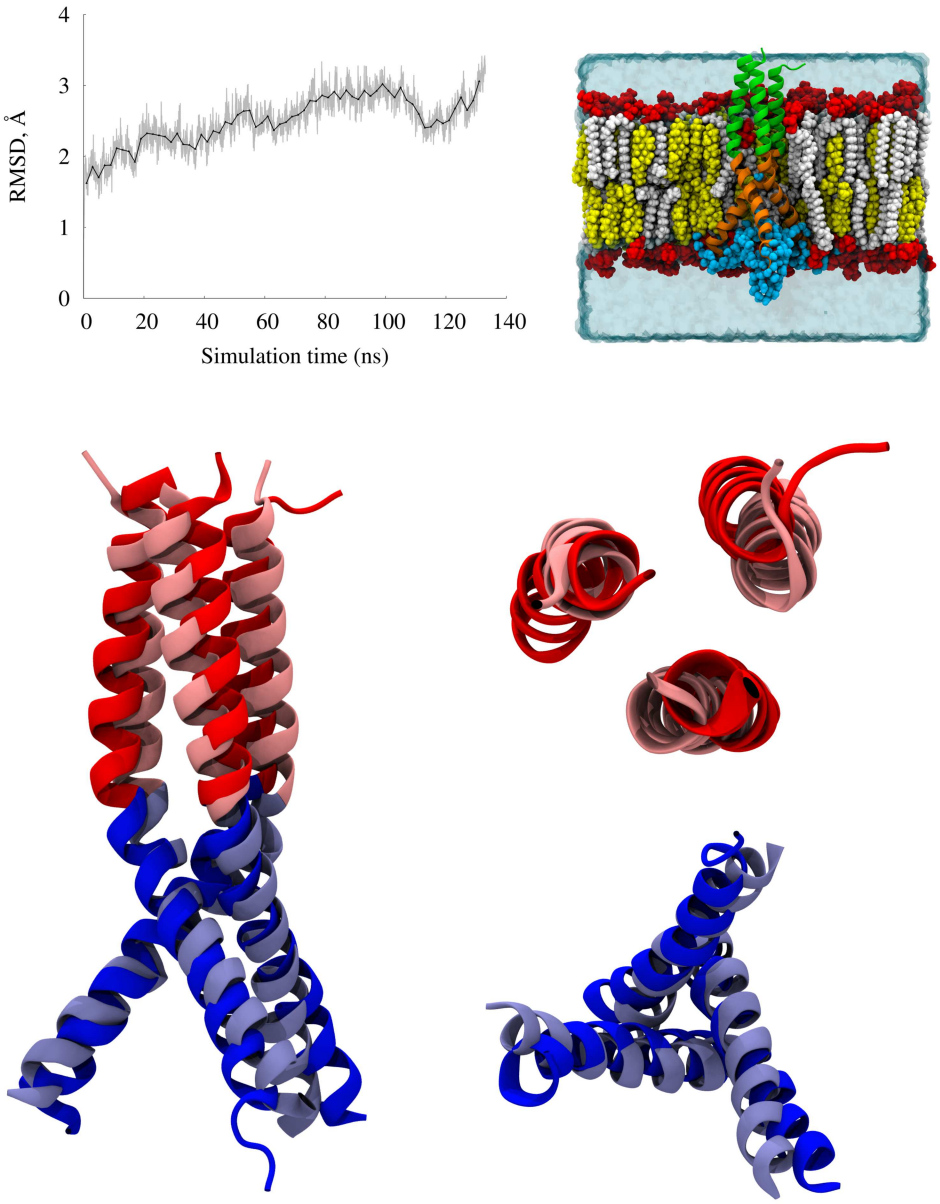

Figure 5.5: Top left, RMSD of backbone of trimeric MPER-MSD vs simulation time in ns. Top right, MPER-MSD trimer model after 133 ns MD: lipid headgroups in red, lipid tails in white, cholesterol in yellow, water in cyan, MSD (residues 682 to 707) in orange, and MPER (residues 660 to 681) in green. Bottom left, MPER-MSD trimer model after 133 ns MD (MPER in red, MSD in blue) compared to vacuum trimer (MPER in light red, MSD in light blue). Bottom right, top view of MPER N-terminus and bottom view of MSD C-terminus.

oxygens show that cholesterol interacted with K681, specifically with its terminal positively charged atoms, NH_3^+ (top panel of Figure 5.7). Other residues both N- and

C-terminal to K681 did not interact as strongly with cholesterol hydroxyl oxygens (bottom three panels of Figure 5.7). This interaction between K681 and cholesterol was unique to this configuration. Similar *g(r)* for shorter trimeric simulations and the monomeric simulations from Chapter 4 are shown in Appendix Figure 5.11. Both the MPER-MSD trimer after 133 ns and the MSD trimer after 150 ns did not have as sharply defined or as high a value of *g(r)* at short distances.

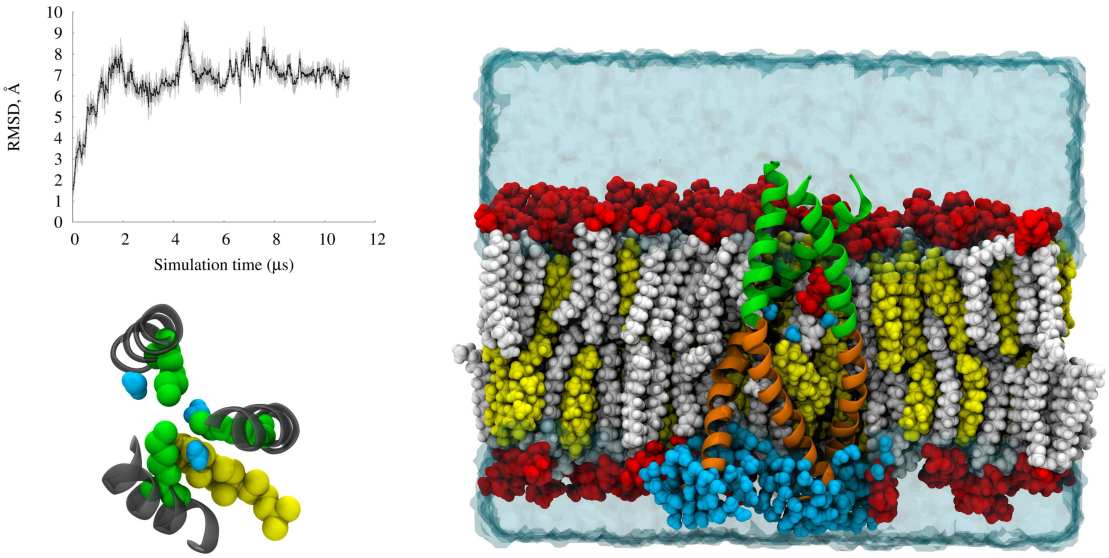

Figure 5.6: Top left, RMSD of backbone atoms vs simulation time for MPER-MSD trimer. Right, MPER-MSD trimer model after 11 μs MD: lipid headgroups in red, lipid tails in white, cholesterol in yellow, water in cyan, MSD (residues 682 to 707) in orange, and MPER (residues 660 to 681) in green. Bottom left, environment near K681, shown from N-terminus, at 11 μs: L681 in green, 3 waters in blue, cholesterol in yellow, and backbone in grey.

This MPER-MSD trimer configuration after 11 μs MD still solvated the three midspan arginines, but did not display any water defects near the midspan arginines. The number of unique water molecules that solvated each residue for each helix, graphed in Appendix Figure 5.12, shows water molecules solvated K681 (as shown

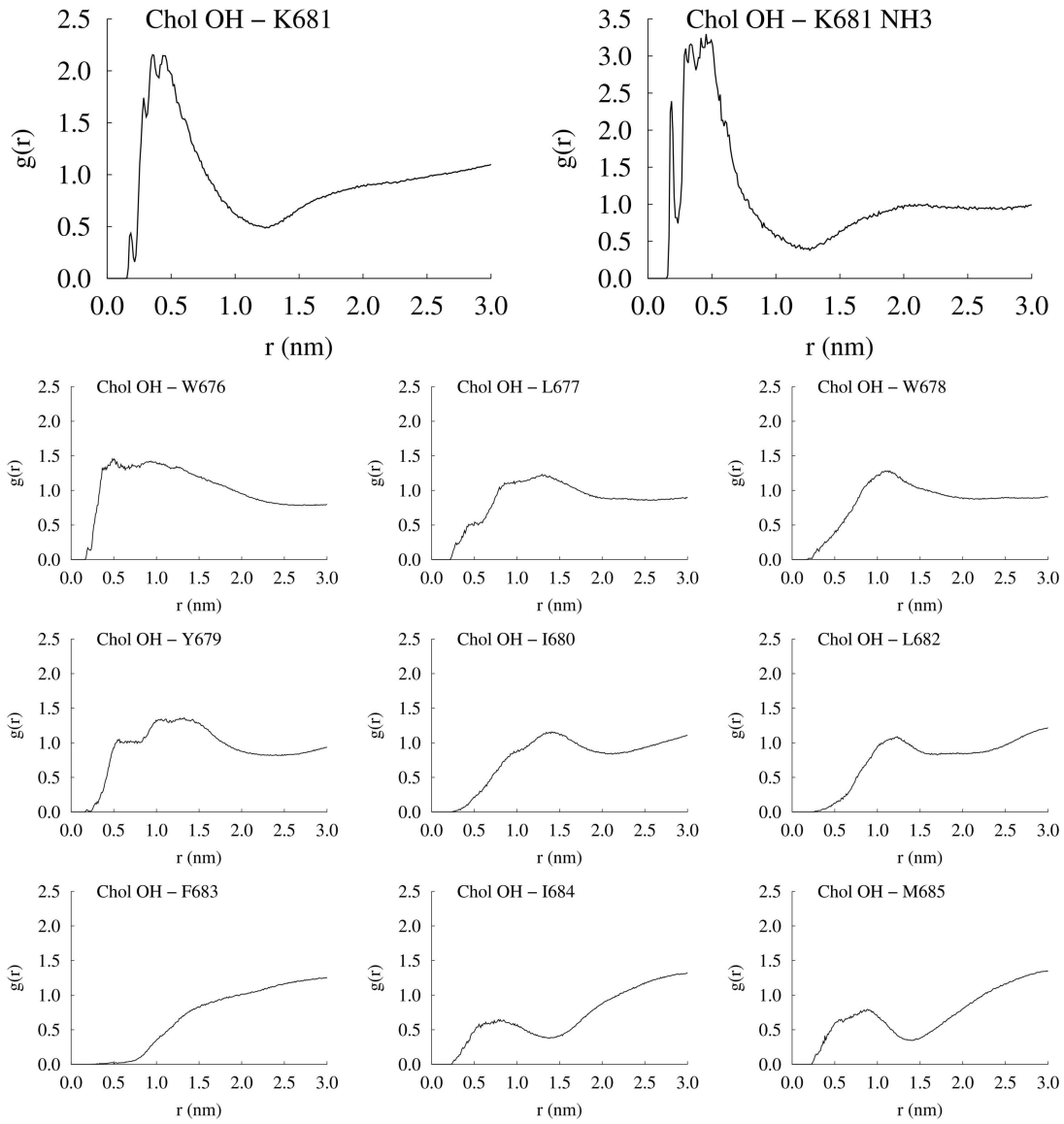

Figure 5.7: Radial distribution function, *g(r)*, vs distance (in nm) between cholesterol hydroxyl oxygen and residues 676 to 685 or residue 681 NH_3^+ atoms.

in Figure 5.6) and R694. However, maps of average minimum water depth along membrane normal show that water did not penetrate the bilayer in the lower leaflet, even at the midspan arginines. Water penetrated the bilayer in the upper leaflet at the trimer, creating an upper leaflet water defect, as seen in Figure 5.8. As before, water penetration near the trimer was strongly fluctuating.

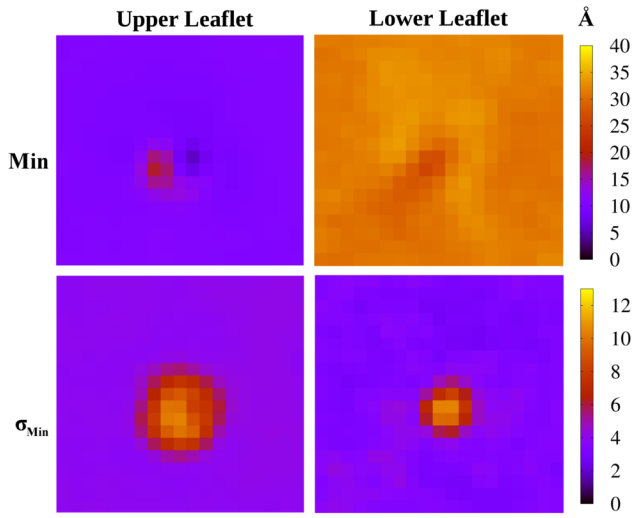

Figure 5.8: Maps of average minimum distance along membrane normal in both leaflets of water molecules from global center of mass of lipid bilayer, Min (top), in Å and standard deviation, σ_{Min}, (bottom) in Å.

Perhaps the absence of the lower leaflet water defects can be attributed to the "sinking" of the trimer in the bilayer, relegating water defects unnecessary. The midspan arginines did not need to snorkel as they were located in the solvated lipid headgroups. Indeed, if the backbone of MPER experiences an upward force at a constant velocity, the water defects quickly reestablished themselves in ∼10 ns, as seen in Figure 5.9. This characteristic would probably be helpful for poration of the membrane by decreasing the energy barrier to fusion. The upwards force may be biologically relevant during the prefusion intermediate stage, when gp41 bridges the cell and viral membranes.

The trimeric MPER structure initially chosen for the MPER-MSD trimer model, as mentioned before, is a putative prefusion trimer crystallized with an isoleucine zipper motif to facilitate the trimerization [25]. The crystal structure agrees qualitatively with the inability of antibodies to access MPER prefusion as the epitope regions are located in the MPER interface. The model of MPER moves away from this struc-

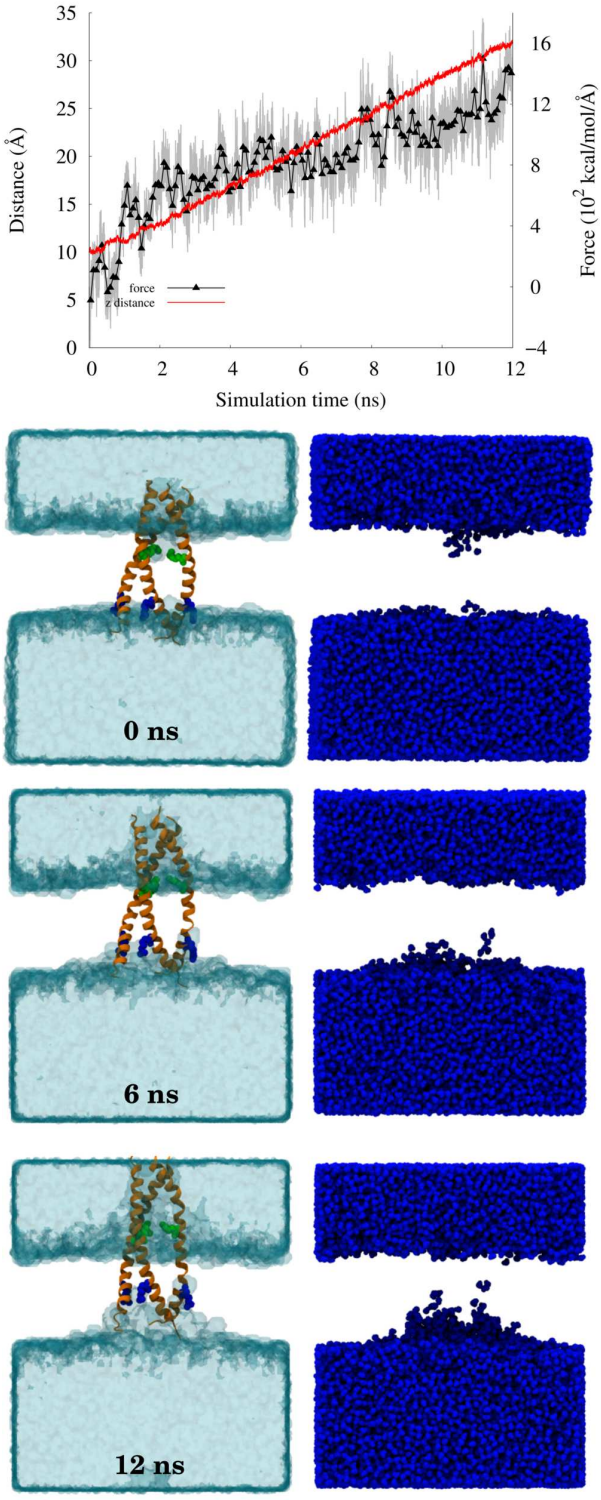

Figure 5.9: Z-distance (y axis) and SMD force (y2 axis) vs simulation time in ns for SMD pulling simulation. Snapshots of trimer at 0, 6, and 12 ns: on left, trimer in orange, R694 in blue, K681 in green, and water in cyan surface; on right, water in blue.

ture, as shown in Figure 5.10. The RMSD of MPER with respect to the initial crystal structure (PDB 3G9R) and to the configuration of the model at timestep 0 of 11 μs MD (after equilibration with constraints and equilibration without constraints for 133 ns) increased within the first μs and then stabilized to within 4 or 5 Å. The MPER model also stabilized to $\sim$5 Å RMSD with respect to a recent trimeric MPER NMR structure that is purported to be a prefusion intermediate structure [137]. This NMR structure (PDB 2LP7), assembled in a DPC micelle, also has its N-terminal residues interacting. Along the C-terminus, the helices increase in distance away from each other. Also, the epitope regions are exposed and more accessible to antibodies, which qualitatively agrees with antibody activity during the prefusion intermediate stage of infection. The MPER-MSD model after 11 μs resembled the NMR structure in that the C-terminal residues of MPER have moved away from each other, exposing the epitope regions, while the N-terminal residues still interacted. The MPER-MSD model was $\sim$5 Å RMSD from the NMR structure. The crystal structure, NMR structure, and model after 11 μs are compared in Figure 5.10.

An interpretation of the relationship between the three trimeric MPER structures is that the trimeric MPER in the MPER-MSD model is an intermediate structure between the prefusion crystal structure (PDB 3G9R) and the prefusion intermediate NMR structure (PDB 2LP7) and would need much longer simulation time to transition to the more open structure. Also, the three structures are reflections of their construction and environment. The crystal structure is aqueous, the NMR structure is encapsulated with a micelle, and the model is exposed to both a cholesterol-containing membrane and solvent. The experimental structures are not restricted by an MSD trimer and the experimental conditions, by necessity, ignore the influence of cholesterol. All three structures lack restrictions from the spike N-terminal to MPER. While the prefusion MPER structure is supported in a metastable state by gp120 and

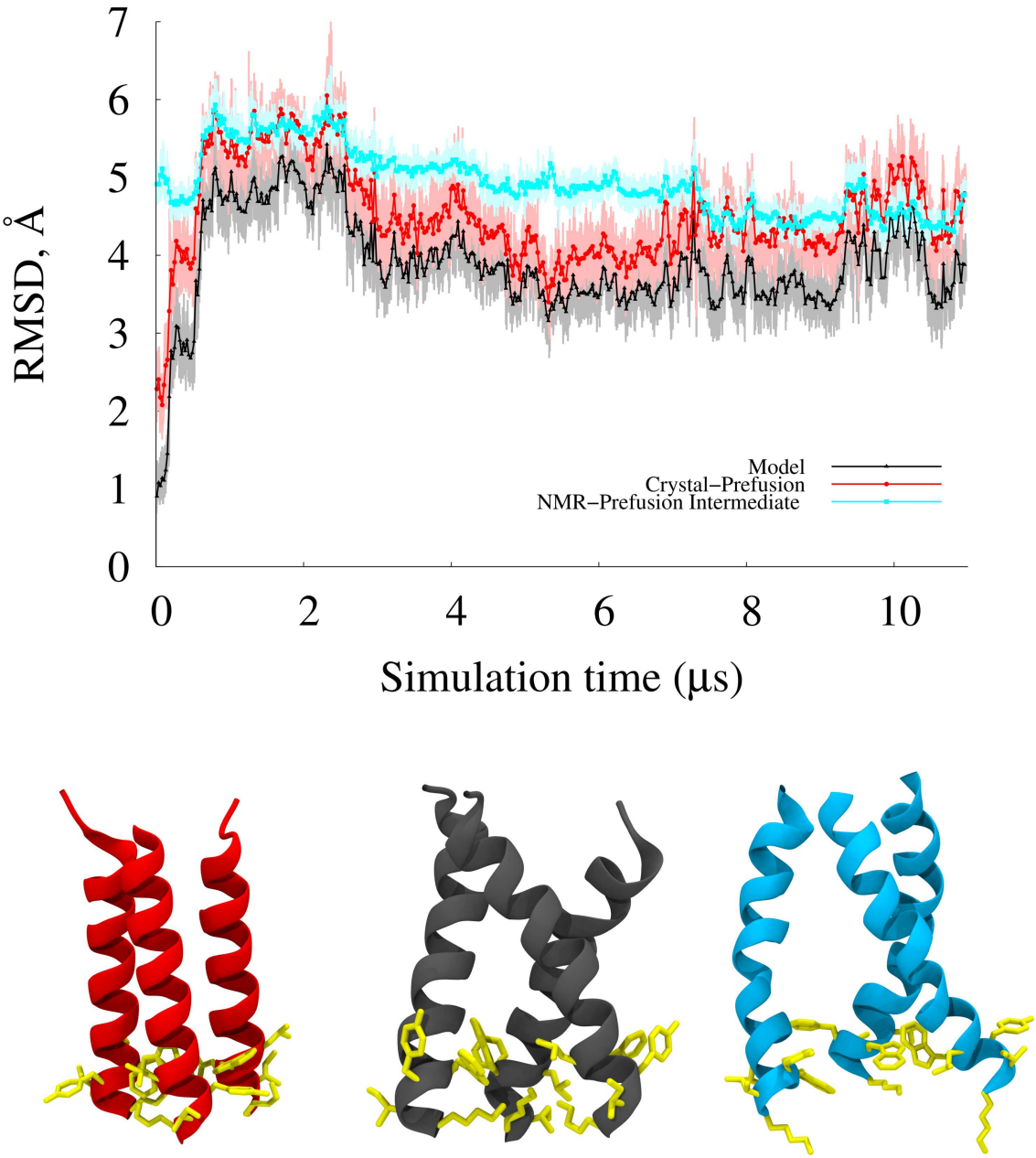

Figure 5.10: Top panel: RMSD of backbone atoms vs simulation time with respect to frame 0 of MPER (black lines), putative prefusion crystal structure of MPER (PDB 3G9R, red lines), and putative prefusion intermediate NMR structure of MPER (PDB 2LP7, blue lines). Bottom panel: MPER backbones shown for comparison, from left to right, PDB 3G9R in red, model after 11 μs in grey, and PDB 2LP7 in blue. Sidechains of CRAC motif rendered in yellow.

perhaps the rest of gp41, the prefusion intermediate MPER structure is on the brink of experiencing strong forces emanating from 6HB formation by the heptad repeat sections. Without further research, it is merely speculation on the correlation between the three MPER trimer structures. However, the trimeric MPER-MSD model is the first all-atom simulation for this region of gp41 and can suggest mechanisms for antibody interaction, cholesterol sequestration, or poration potential from upper and lower leaflet water defects.

5.3.5 Summary and conclusions of Aim III

An all-atom model of trimeric MPER-MSD was constructed and equilibrated in a cholesterol-containing bilayer, representing the first such simulation in literature. Initially, the lowest energy MSD trimer identified in vacuum was equilibrated in a bilayer, but required trimeric MPER, excised from a putative prefusion crystal structure, to stabilize. Two orders of magnitude longer MD (to 11 μs) culminated in the MPER-MSD model transitioning to a different, stable configuration that qualitatively resembled a recent prefusion intermediate NMR structure. The model was almost completely immersed in the cholesterol-containing bilayer, but the helices near the CRAC motif separated to allow sequestration of a cholesterol molecule and water. The model potentially shows a transition structure between the prefusion and prefusion intermediate stages of HIV-1 fusion and infection.

5.3.6 Appendix

Table 5.2: Comparison of HIV-1 gp41 MSD peptide sequences

HxB2	K	**L**	F	I	M	I	**V**	G	G	L	**V**	G	L	R	I	V	F	**A**	V	L	S	I	**V**	N	R	V	R
BaL	K	**I**	F	I	M	I	**I**	G	G	L	**I**	G	L	R	I	V	F	**S**	V	L	S	I	**M**	N	R	V	R

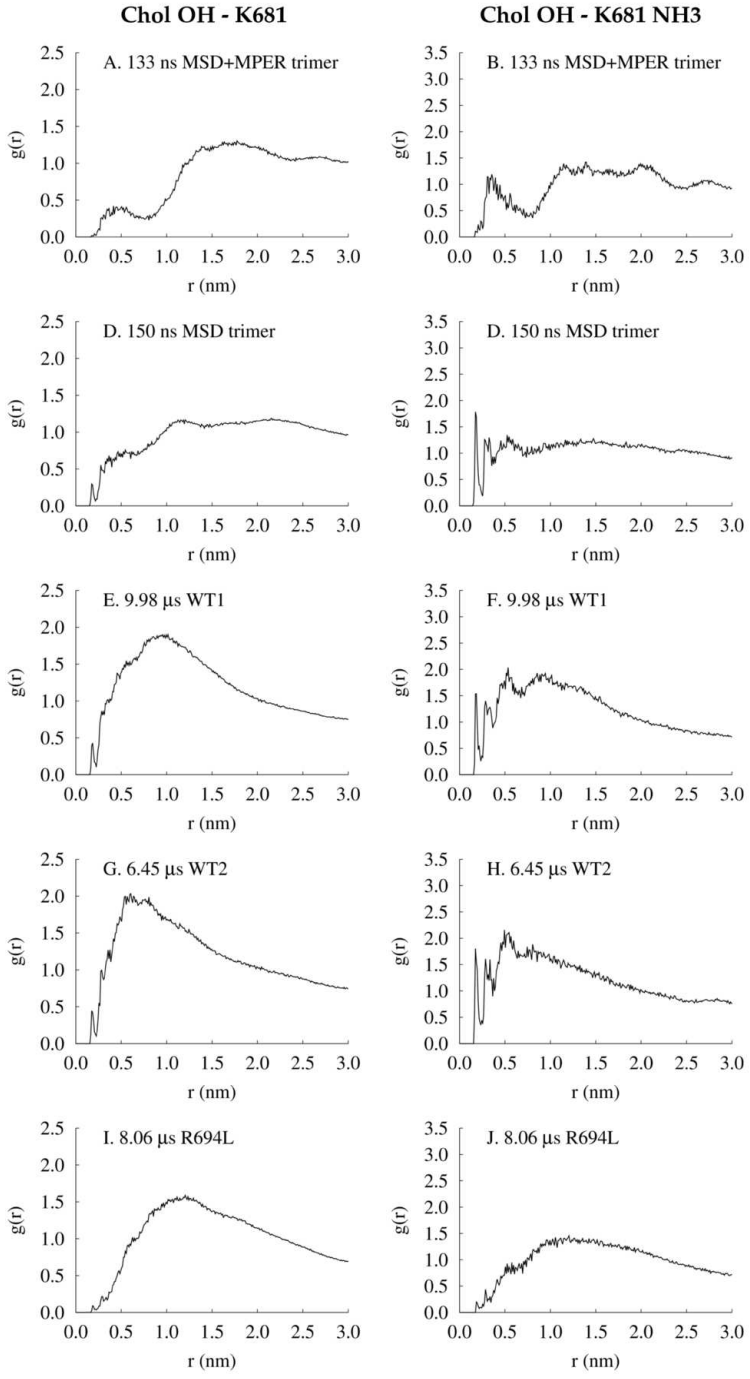

Figure 5.11: Radial distribution function (*g(r)*) vs distance (in nm) between cholesterol hydroxyl oxygen and residue 681 (left panel) or residue 681 NH_3^+ atoms (right panel) for previous systems A, B) 133 ns MPER-MSD trimer, C, D) 150 ns MSD trimer, E, F) WT1, G, H) WT2, and I, J) R694L.

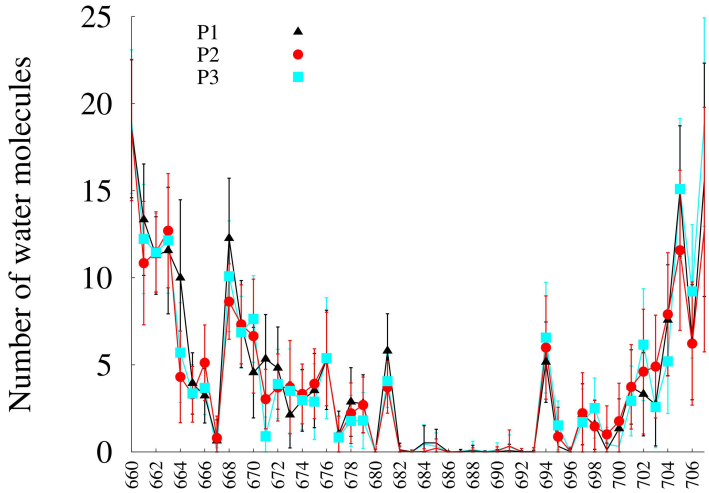

Figure 5.12: Number of water molecules within 4 Å of protein vs amino acid that each water molecule is uniquely attributed to during equilibrium MD. All statistics are from the full trajectory and error bars represent standard deviation.

6. Aim IV: Role of the dimerization motif, GXXXG, in the MSD of trimeric HIV-1 gp41

6.1 Introduction to Aim IV

The GXXXG motif has previously been identified as a transmembrane, α-helical dimer interaction motif for right-handed helices either as dimers or in complex, multi-protein structures in the bilayer [48, 138, 139]. HIV-1 MSD has this conserved motif, where the first X is also a glycine (GGXXG). It has been speculated that the motif is important for MSD trimerization, either in the native structure or during fusion. Experiments suggest a more complex role though. Replacement of the MSD with that of CD4, which also contains GGXXG, did not affect fusion [140]. Replacement of the MSD with glycophorin A (GpA), a well known transmembrane dimer with the GXXXG motif, impaired fusion; function was not restored with mutation to the dimerization motif [46]. The dimerization or trimerization ability of the motif is not sufficient for MSD function and perhaps the residues surrounding the motif may modulate the affinity of interhelical interactions [138]. Mutants of the MSD GXXXG had normal expression, processing kinetics, membrane incorporation, and surface expression, indicating that perhaps the GXXXG motif is not important for expression, but for fusion. In fact, mutations to one or more of the glycines decreased the fusion activity of HIV [51, 52].

Although the lowest energy WT MSD trimer in vacuum from Aim III had interacting GXXXG residues, the motif did not sustain trimerization of excised MSD transmembrane peptides on the 100 ns timescale in MD simulations. Aim IV investigated dimers of the WT MSD on similar timescales in order to assess if the HIV-1 MSD sequence is capable of dimerization. We aimed to answer the following

questions:

1. Does the GXXXG motif sustain dimerization of the WT dimer on the 100 ns timescale?

2. How does the midspan arginine affect dimerization?

3. What do these simulations reveal about the probable oligomeric state of the HIV-1 MSD?

6.2 Methods

6.2.1 Enumeration of MSD dimer

The enumeration of Euler configurations for the WT and R694L dimer used the same procedure as described in Chapter 5, except the peptides were rotated twofold symmetric around the z-axis and not threefold symmetric before translation. These systems were referred to as WT-LED and R694L-LED.

6.2.2 Setup and equilibration of MSD dimer

The lowest energy MSD dimer, WT-LED, was generated by branching off the SMD simulation described in Section 5.2.2 after the 2^{nd} peptide was pulled into the bilayer. The remaining peptide above the bilayer was deleted. The starting structure for the dimer simulation now consisted of two monomeric MSD peptides in a bilayer; this structure was equilibrated for 5 ns. Then, 5 ns of TMD was used to assemble the dimer in the membrane. The dimer backbone was constrained for 50 ns while lipids and water relaxed around it. An interfering cholesterol and lipid molecule were deleted and the constraints were continued for 50 ns. The constraints were released and the dimer was simulated for 150 ns. This system had 64,205 atoms, including

ions, 151 DPPC molecules, 183 cholesterol molecules, 916 protein atoms, and ∼10,000 water molecules. The system size was 90 x 80 x 80 $Å^3$.

A low energy WT dimer was also constructed in the bilayer and equilibrated, similarly to WT-LED, as a different initial condition. This system was referred to as WT-D.

6.2.3 Setup and equilibration of MSD mutant dimers

A R694L dimer was also enumerated to find the lowest energy mutant dimer. A system with the R694L dimer, known as R694L-LED, was prepared similarly to WT-LED by starting with the same initial structure as above, before TMD. The two peptides in random locations were mutated and the entire system was reneutralized, ionized, and minimized before 5 ns of TMD. The dimer was also constrained for 100 ns and then released for 150 ns of production MD.

Another mutant was created by branching off the WT-D system after 50 ns of backbone constraints, mutation of the midspan arginines to leucines, reneutralization, minimization, and simulation with protein constraints for 50 ns. The new system, known as R694L-D, was then simulated for 150 ns production run.

6.2.4 Calculations of observables

The radial distribution function was calculated using the *g(r)* plugin in VMD and a δ of 0.1 Å [117]. The pair selections were the two sidechain hydrogens (HA1 and HA2) of one of the glycines in GGXXG from one helix to another. The number of unique water molecules per residue was calculated similarly to Chapter 3. The NAMD parameters for the dimer simulations were the same as in Chapter 3.

6.3 Aim IV results

6.3.1 Lowest energy structure of MSD dimer in vacuum is determined from ∼270,000 possible geometric configurations

The enumeration of WT and R694L dimers resulted in lowest energy structures that are ∼-10 kcal/mol lower in energy per dimer than the lowest energy trimer structures. WT-LED and R694L-LED therefore had the lowest energy per structure, -57.8 and -58.0 kcal/mol, respectively, compared to -55.8 for the WT trimer. Along with WT-LED and R694L-LED, the Euler transformations are shown for a low energy dimer, WT-D, in Table 6.1. Similarities existed between WT-LED, R694L-LED, WT-D, and the vacuum trimers with an endoplasmically snorkeling arginine. All were left-handed with interacting GXXXG motifs. This is not surprising for the dimers, as GXXXG is a known dimerization motif.

Table 6.1: Lowest energy Euler transformations for dimer enumeration

Monomer	**Energy per dimer (kcal/mol)**	**Energy per structure (kcal/mol)**	$\alpha(^\circ)$	$\beta(^\circ)$	$\gamma(^\circ)$	**Equilateral distance (Å)**	**Handedness**
WT-LED	-28.9	-57.8	30	-42.3	30	7	L
WT-D	-19.5	-39.0	0	-42.3	40	5	L
R694L-LED	-29.0	-58.0	30	-42.3	30	7	L

Conversely, the majority of structurally known transmembrane dimers with the GXXXGXXXG glycine zipper are right-handed and the rest are also predicted to be right-handed [139]. However, a recent 6 Å cryo-EM structure of the uncleaved, unliganded spike finds the α-helical MSD region to be a left-handed trimer with a

crossing angle of 35°, although the MSD cryo-EM region might match the ectodomain region of an earlier cleaved spike cryo-EM [19, 136, 141]. The crystal structure of the MPER trimer with an isoleucine zipper utilized for the model in Chapter 5 and an NMR structure of the MPER trimer with a bacterial trimerization domain in a DPC micelle are both left-handed [25, 137]. Also, the 6HB post-fusion structure is left-handed [21]. The MSD may be an exception among right-handed transmembrane α-helices with a glycine motif; the HIV-1 experimental studies cited above support a left-handed trimer for the MSD.

6.3.2 Production runs of dimer were stable

The lowest energy dimer, WT-LED, and a low energy dimer, WT-D, were both simulated for 150 ns in a cholesterol-containing bilayer and both reached RMSDs of $\sim$4 Å with respect to their vacuum structure, as shown in Figure 6.1A. The dimers remained α-helical and membrane spanning for the duration of the trajectories. The increase in RMSD over the trajectory is partly a result of the decrease in tilt angles of both helices of WT-D and WT-LED, as seen in Figure 6.1B, although the GXXXG motifs remained associated. The decrease in tilt angles can be seen in the snapshots of Figure 6.1, where the light blue dimers are the vacuum structures and the dark blue dimers are the structures after 150 ns MD for WT-LED and WT-D. Both dimers show associated GGXXG residues with decreases in tilt angles.

The water defects associated with the midspan arginines were similar to the water defects explored in Aim I and the number of unique water molecules per residue for both helices are shown in Figure 6.2. WT-LED has 4.1 $\pm$ 1.2 and 5.9 $\pm$ 1.4 (SD) unique water molecules at the midspan arginines while WT-D has 4.6 $\pm$ 1.4 and 5.7 $\pm$ 1.3 (SD). This compares very well with the average of 4.8 unique water molecules at the midspan arginines of the monomeric MSDs in Aim I.

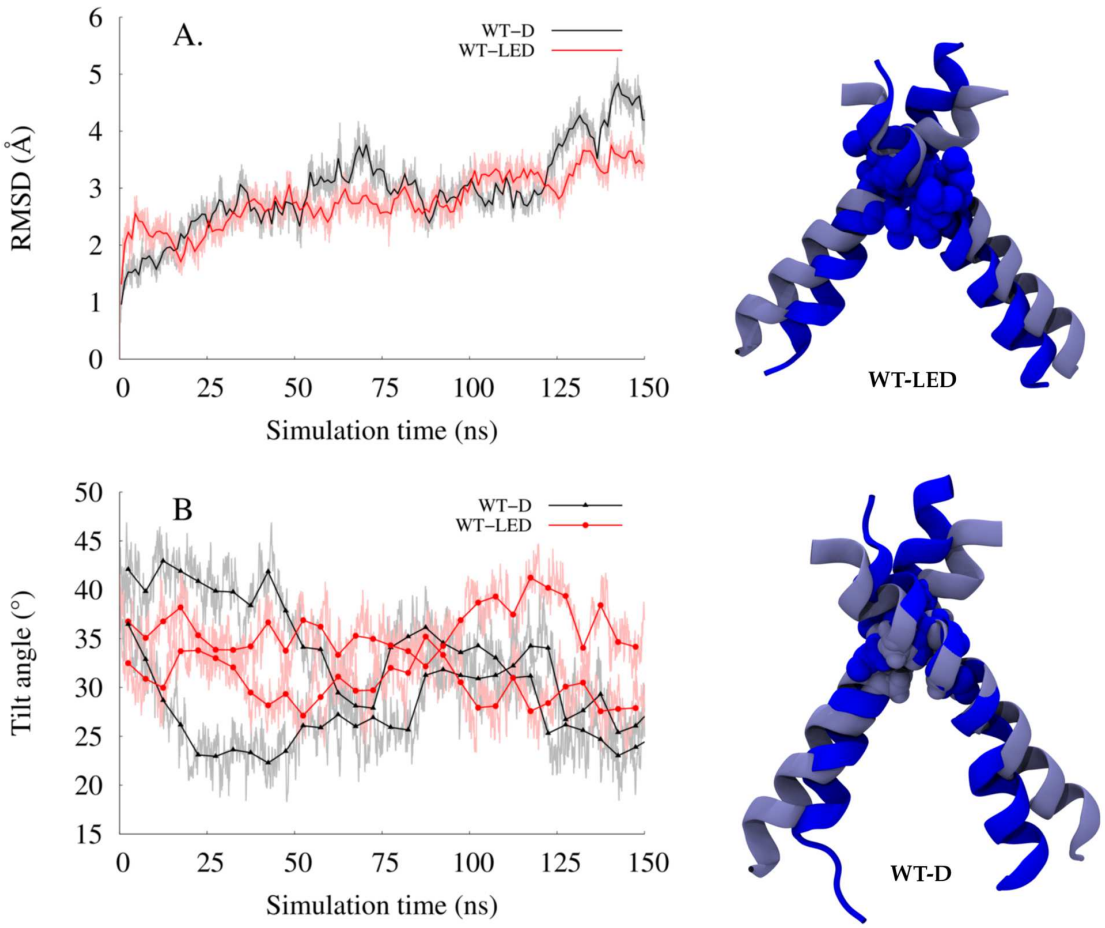

Figure 6.1: A) RMSD of WT-LED and WT-D backbone vs simulation time. B) Tilt angle of all helices with respect to membrane normal vs simulation time. Snapshots: initial structure in light blue and structure after 150 ns in dark blue for WT-LED and WT-D. GGXXG residues in vdW.

The lowest energy mutant dimer, R694L-LED, and a mutant dimer, R694L-D, also remained α-helical and membrane spanning during the 150 ns trajectories. The RMSD with respect to vacuum structure vs simulation time is shown in Figure 6.3A for both dimers. Both R694L-LED and R694L-D reached $\sim$4 Å RMSD during the trajectories. Similarly to the WT dimers, the mutant dimers had decreases in tilt angle of both helices, as shown in Panel B and accompanying snapshots. The thickness of the membrane probably causes the helix tilts to decrease in all systems, as evidently there is a hydrophobic mismatch initially between the vacuum structures and the

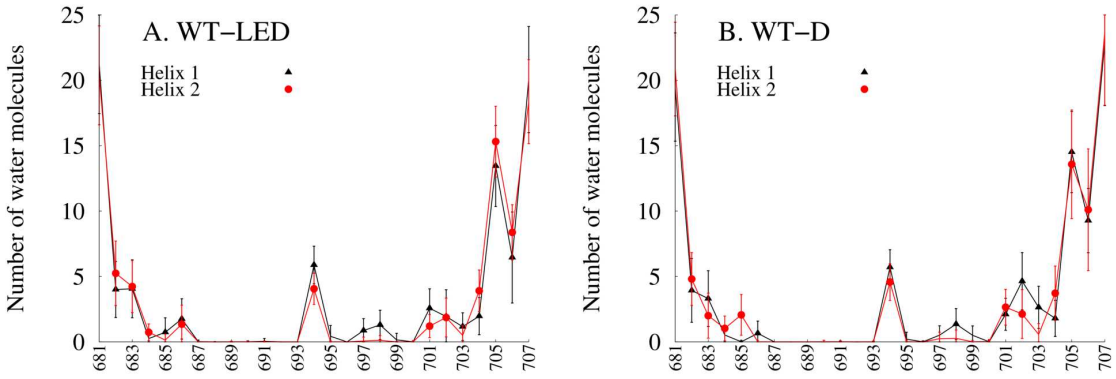

Figure 6.2: Number of unique waters per residue for each helix of A) WT-LED and B) WT-D, with standard deviation.

membrane. Decrease of tilt angle was also seen in Aim III with the 11 μs simulation of the MPER-MSD trimer. The lowest energy vacuum structures of the dimers and trimers all have decreased tilt angles after simulation in a membrane that approach the tilt angles of the monomeric MSDs in Aim I.

6.3.3 Interactions of GXXXG residues differed between MSD dimers and trimers

Although the dimers reached ∼4 Å RMSD from their initial configurations, the GXXXG motifs remained interacting. This can be seen more clearly with graphs of radial distribution function over time for the dimers. Graphed in the Appendix are *g(r)* vs distance in nm for atomic pairs of sidechain hydrogens of either residue 688, 689, 692 in helix 1 (H1) with respect to sidechain hydrogens of either residues 688, 689, 692 in helix 2 (H2), resulting in 9 potential interactions for WT-LED, WT-D, R694L-LED, and R694L-D. The *g(r)* for the first 50 ns, middle 50 ns, and last 50 ns are shown as three lines in each graph (black, red, blue, respectively). The first thing of note is that the *g(r)* profiles were not similar for WT-LED and WT-D or for R694L-LED and R694L-D. For instance, WT-D had the greatest probability

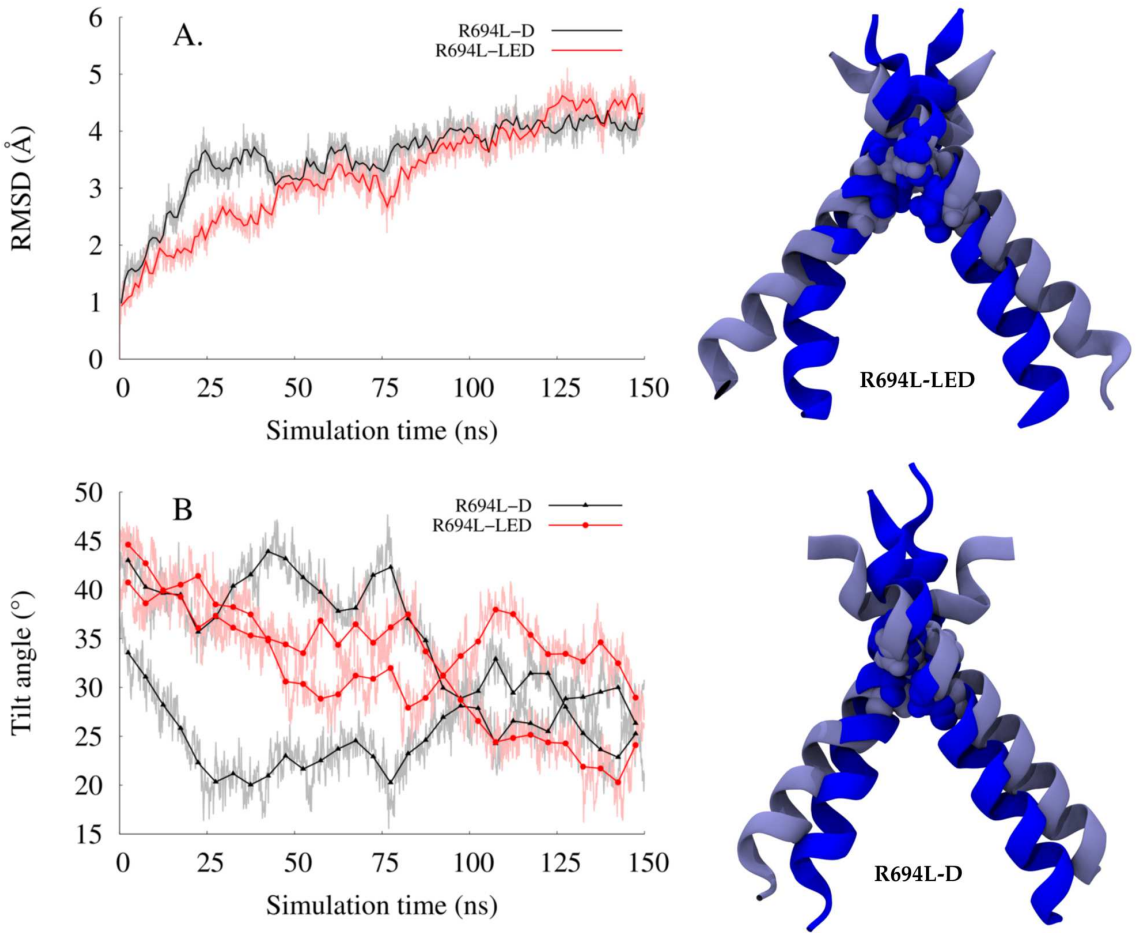

Figure 6.3: A) RMSD of R694L-LED and R694L-D backbone vs simulation time. B) Tilt angle of all helices with respect to membrane normal vs simulation time. Snapshots: initial structure in light blue and structure after 150 ns in dark blue for R694L-LED and R694L-D. GGXXG residues in vdW.

for the pair of hydrogens from 689 in H1 to 688 in H2 and this profile was constant with time, with a 1^{st} peak at ∼3 Å. The same graph for WT-LED also had a peak at ∼3 Å, but the probability was half that of WT-D and did vary with time. For R694L-LED, its greatest probability was for residue 689 in H1 to residue 689 in H2. However, the 1^{st} peak decreased from 4.2 Å to 2.3 Å over the simulation, indicating the atoms moved closer. For R694L-D, this interaction was constant and the 1^{st} peak remained at 3.0 Å over the trajectory. Overall, the dimers have strong interactions

between GGXXG residues, but these interactions are not the same between systems and sometimes change with time. Also, the midspan arginine did not have an effect on the dimerization ability of the GGXXG motif, as the mutant dimers had as strong and as close interactions as the WT dimers.

Also graphed in the Appendix are the same set of *g(r)* graphs for the trimer simulations from Aim III. The MSD trimer simulated for 150 ns shows much smaller probability than the dimers and with 1^{st} peaks greater than 5 Å. This is consistent with the conclusion in Aim III that the GGXXG motif was not able to sustain trimerization on the 150 ns MD timescale. In contrast, the MPER-MSD trimer of Aim III was able to sustain trimerization on the same timescale and the graphs reflect that. The greatest interaction was with residue 688 of H1 to residues 688 in H2 and H3, with a constant 1^{st} peak at 3.7 Å over time. As discussed previously, these interactions between GGXXG residues do not last at longer timescales, as the MPER-MSD system that was simulated for 11 μs has extremely small *g(r)* and at larger distances than the other systems.

Therefore, the *g(r)* profiles show that the dimers (both WT and mutant) can sustain dimerization on the 150 ns timescale, while the trimers cannot. (Further simulations, up to 10 μs, would be able to assess these dimeric interactions on a longer timescale.) However, interactions between GGXXG residues seem to be nonspecific as the profiles differed among the dimers and their replicates. The extra glycine in MSD compared to the GXXXG motif may account for this flexibility. It is unknown why gp41 or spike might need a stronger dimerization than trimerization ability at the MSD. The ability of MSD to dimerize may be necessary during fusion for interactions with other spikes or asymmetric 6HB formation [142]. Another instance may exist after translation at the ER [7, 143]. Env proteins gp120 and gp41 are translated in the ER as monomeric, covalently-linked gp160. Most gp160s trimerize, but other

oligomeric forms are found. Dimeric gp160 may be a prerequisite for trimeric gp160, as sequential oligomerization steps may be more plausible than concurrent trimerization. Trimeric gp160 is then trafficked to the Golgi, where it is proteolytically cleaved.

6.3.4 Summary and conclusions of Aim IV

Previously, the presence of the GXXXG motif in the MSD (as GGXXG) was not sufficient for trimerization, as the lowest energy MSD trimer in vacuum did not remain trimerized in the bilayer on the 100 ns timescale in Aim III. Also in Aim III, the addition of MPER stabilized the GGXXG-mediated trimerization on the 100 ns, but not the 10 μs, timescale. However, Aim IV showed that the lowest energy WT dimer (WT-LED) in vacuum remained dimerized, with persistent GGXXG interactions as seen with *g(r)* over time. The lowest energy R694L dimer, R694L-LED, also remained associated on the 100 ns timescale. So did other simulations with different initial conditions of the WT and R694L dimers (WT-D and R694L-D). This indicates that either the GGXXG residues have a more complex role in HIV fusion than mediating trimerization or that the motif is required for dimerization. The MSDs may dimerize during the HIV-1 lifecycle, such as oligomerization after translation at the ER, as asymmetric 6HB formation, or as needed for interactions with other spikes during fusion. Since the native structure of gp41 is metastable, a native MSD trimer may be as well and require more or all of gp41 to stabilize with GGXXG interactions. A simulation of the dimer for many μs can further indicate if the GGXXG residues mediate dimerization, not trimerization, of the MSD or if there is something specific about the MSD sequence that resists oligomerization, such as the midspan arginine and associated water defect.

6.3.5 Appendix

The following graphs show *g(r)* vs distance (nm) for the first, middle, and last segments of the simulations (black, red, blue, respectively). The interaction pairs for the dimers are the sidechain hydrogens of either residue 688, 689, 692 in helix 1 (H1) with respect to sidechain hydrogens of either residues 688, 689, 692 in helix 2 (H2), resulting in 9 potential interactions for WT-LED, WT-D, R694L-LED, and R694L-D. For the trimers from Aim III simulated on the 100 ns timescale, the interaction pairs are the sidechain hydrogens of either residue 688, 689, 692 in helix 1 (H1) with respect to sidechain hydrogens of either residues 688, 689, 692 in helices 2 and 3 (H2 and H3). This was because interactions from H1 to H2, H2 to H3, and H1 to H3 were similar to the pair H1 and H2/H3. This was not the case for the 11 μs MPER-MSD trimer model, so the *g(r)* graphs for this simulation are for the pairs H1 to H2, H2 to H3, and H1 to H3 in the final 3 figures. Note that the scale is not the same between systems.

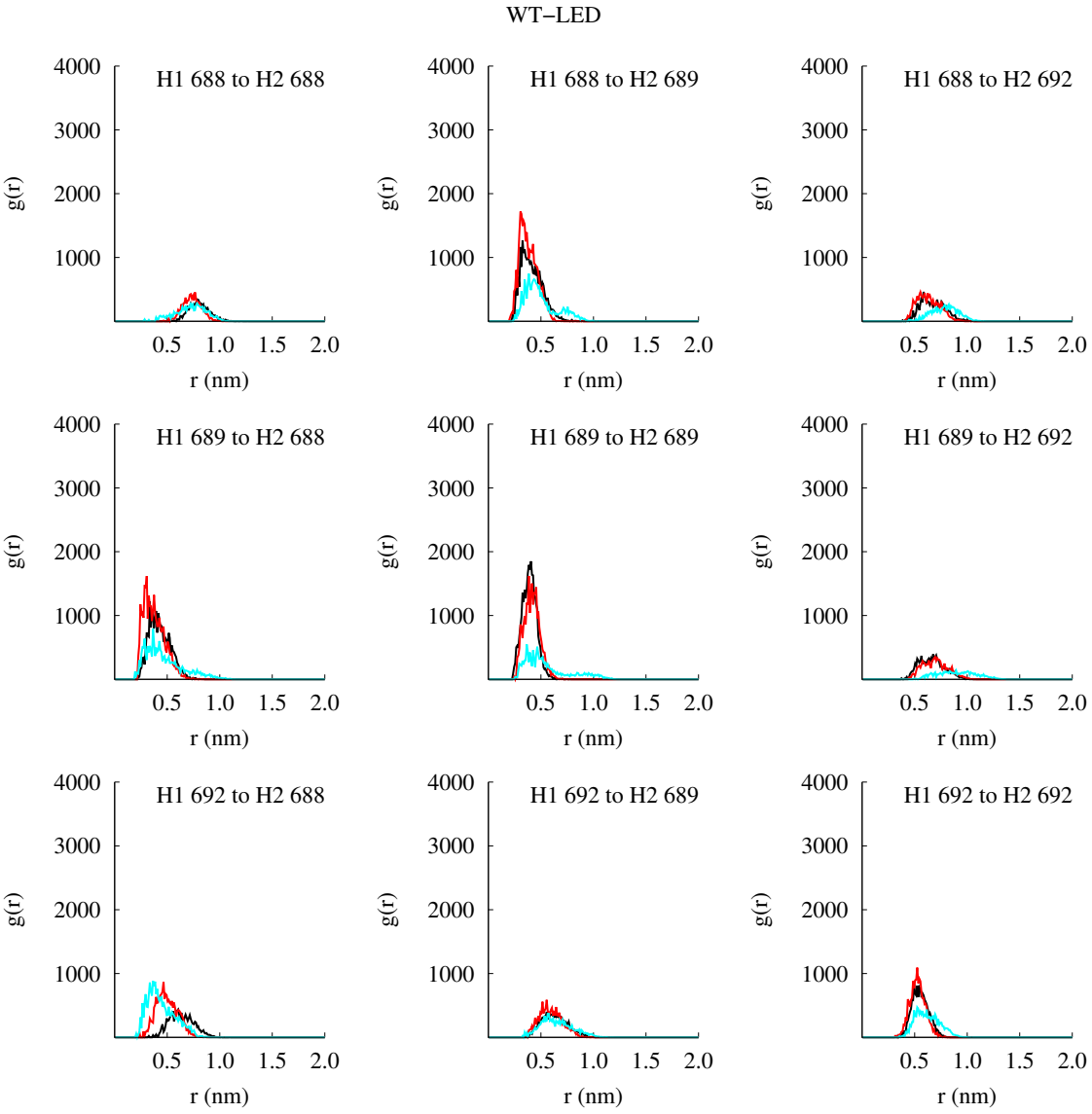

Figure 6.4: Radial distribution function vs distance in nm for WT-LED. 0-50 ns in black, 50-100 ns in red, 100-150 ns in blue.

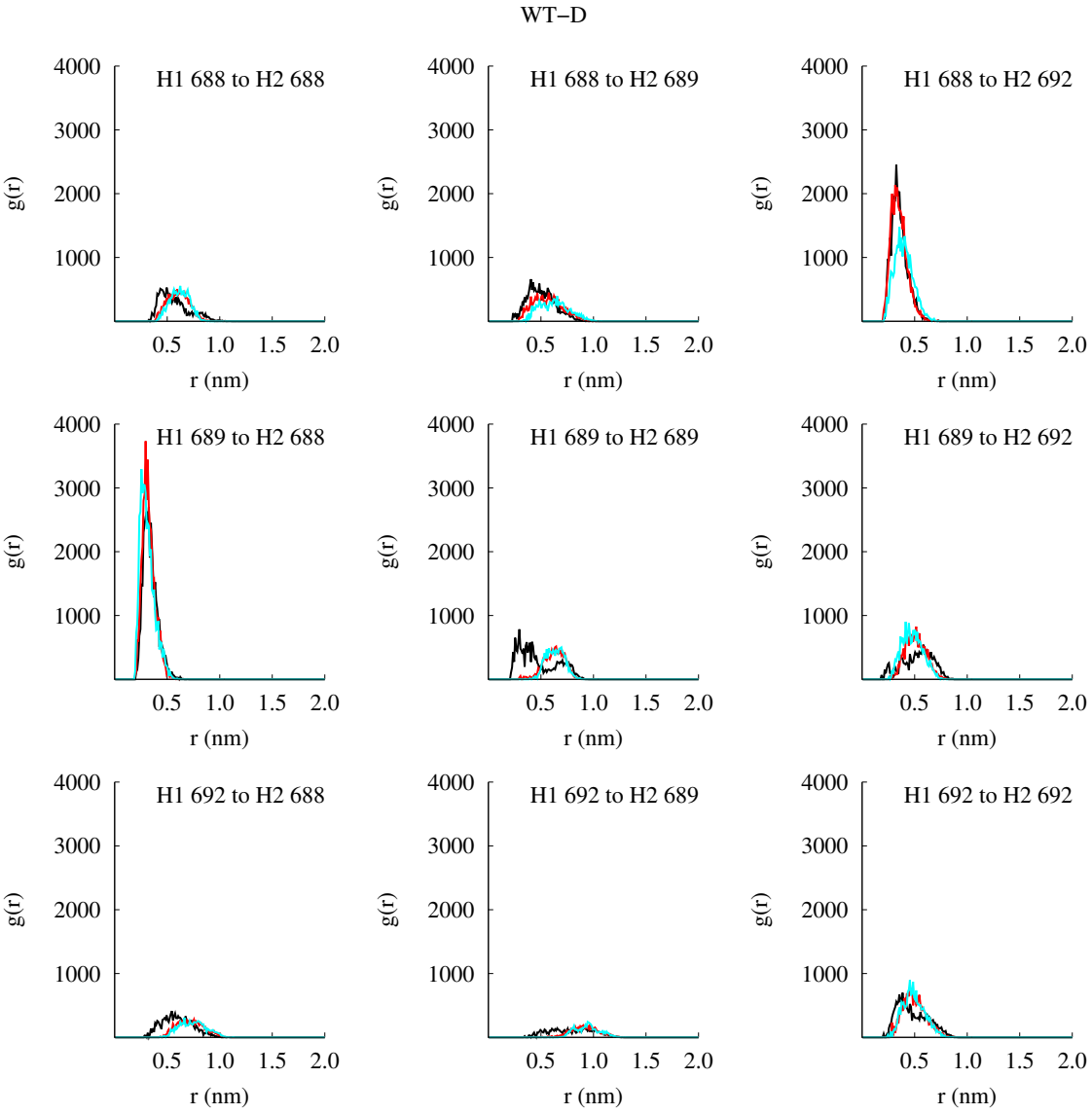

Figure 6.5: Radial distribution function vs distance in nm for WT-D. 0-50 ns in black, 50-100 ns in red, 100-150 ns in blue.

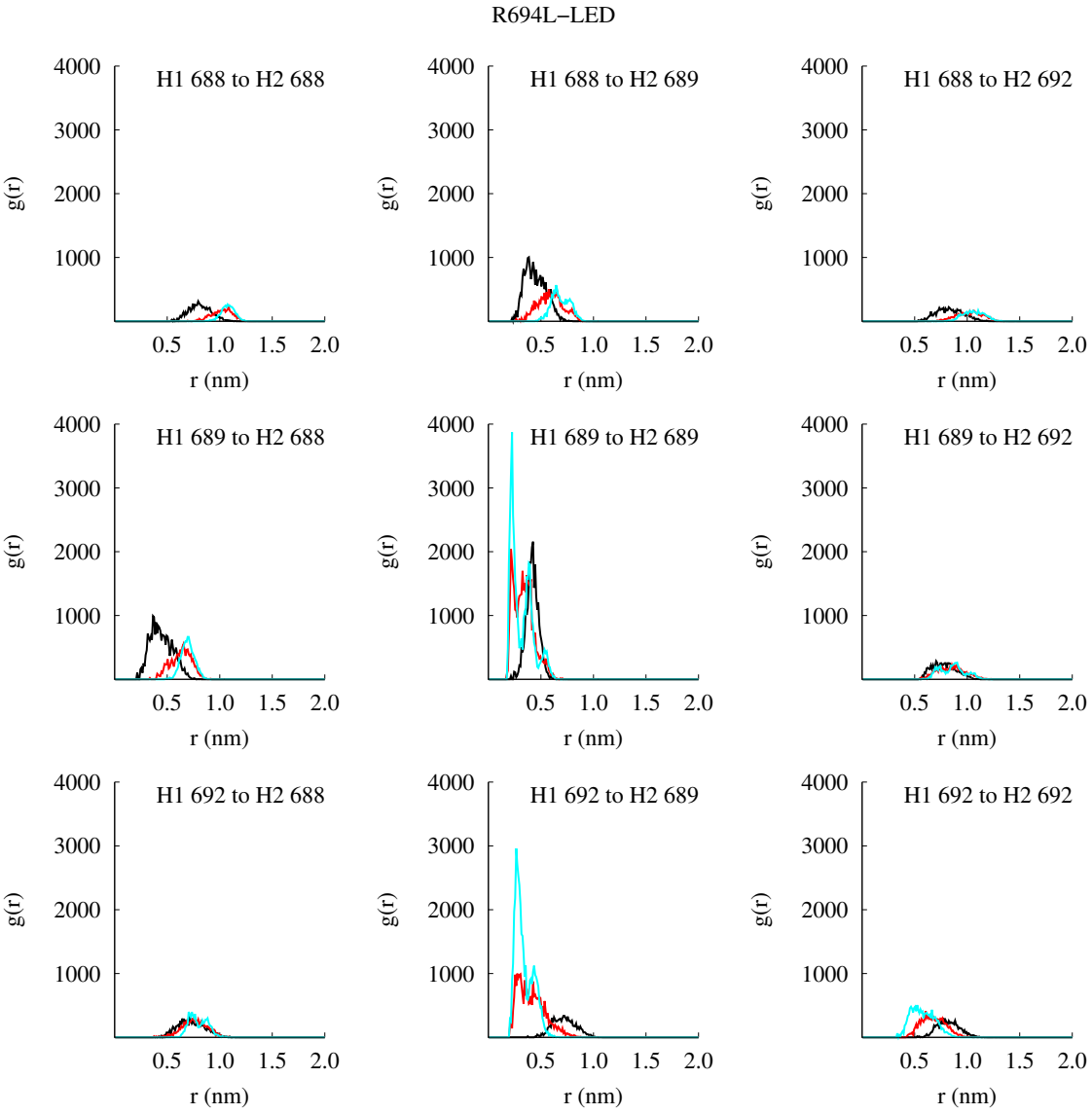

Figure 6.6: Radial distribution function vs distance in nm for R694L-LED. 0-50 ns in black, 50-100 ns in red, 100-150 ns in blue.

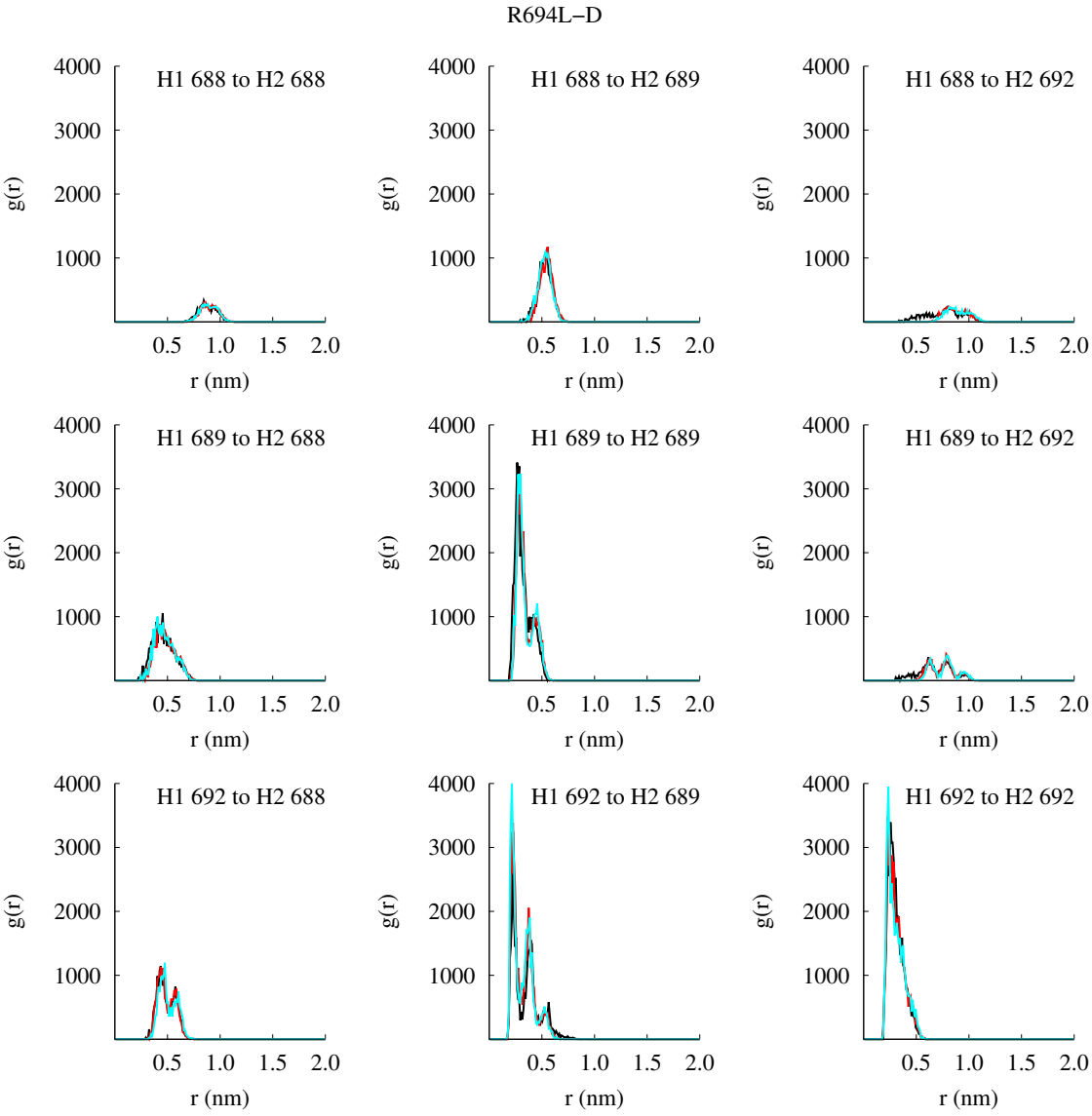

Figure 6.7: Radial distribution function vs distance in nm for R694L-D. 0-50 ns in black, 50-100 ns in red, 100-150 ns in blue.

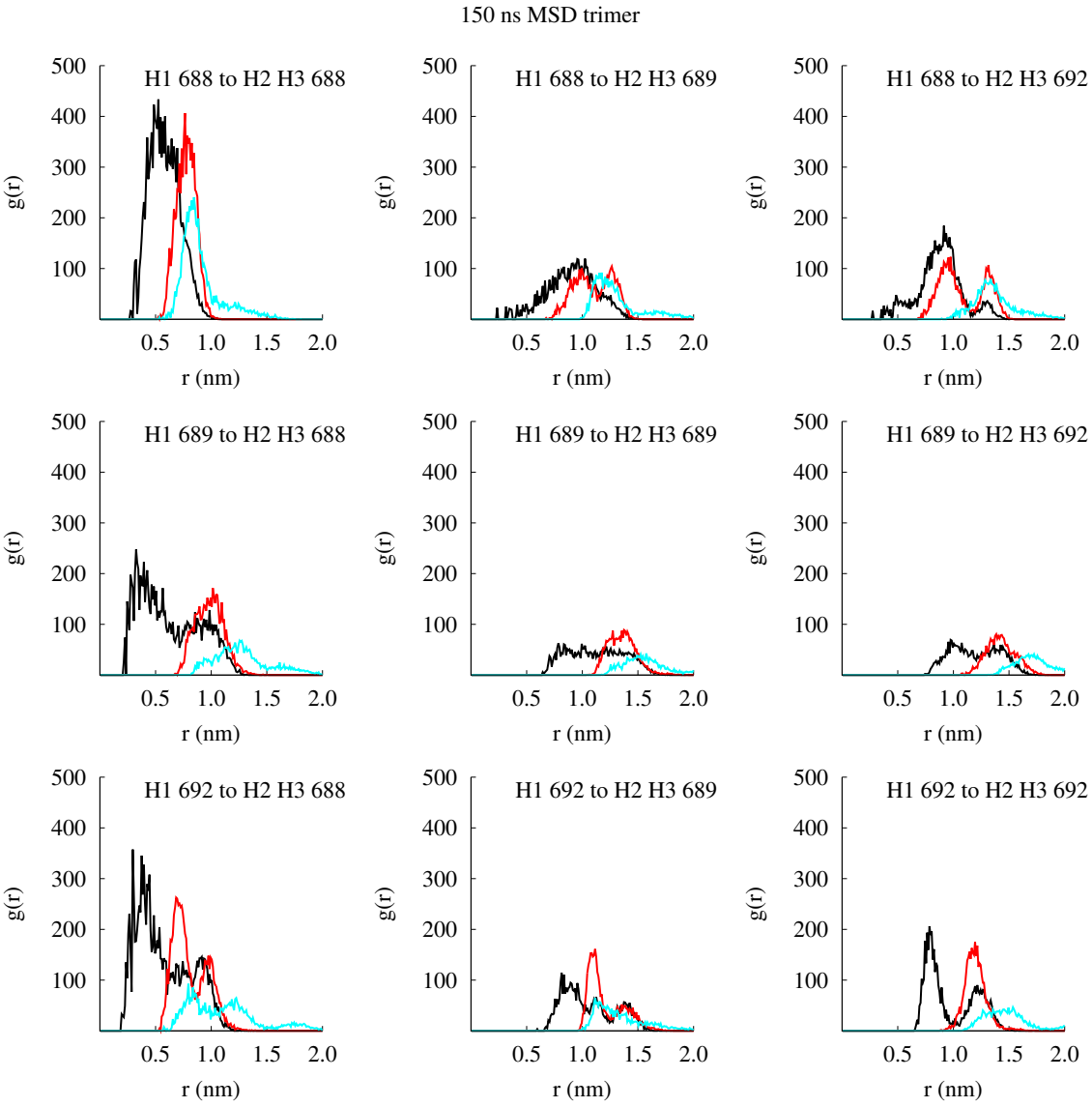

Figure 6.8: Radial distribution function vs distance in nm for 150 ns MSD trimer from Chapter 5. 0-50 ns in black, 50-100 ns in red, 100-150 ns in blue.

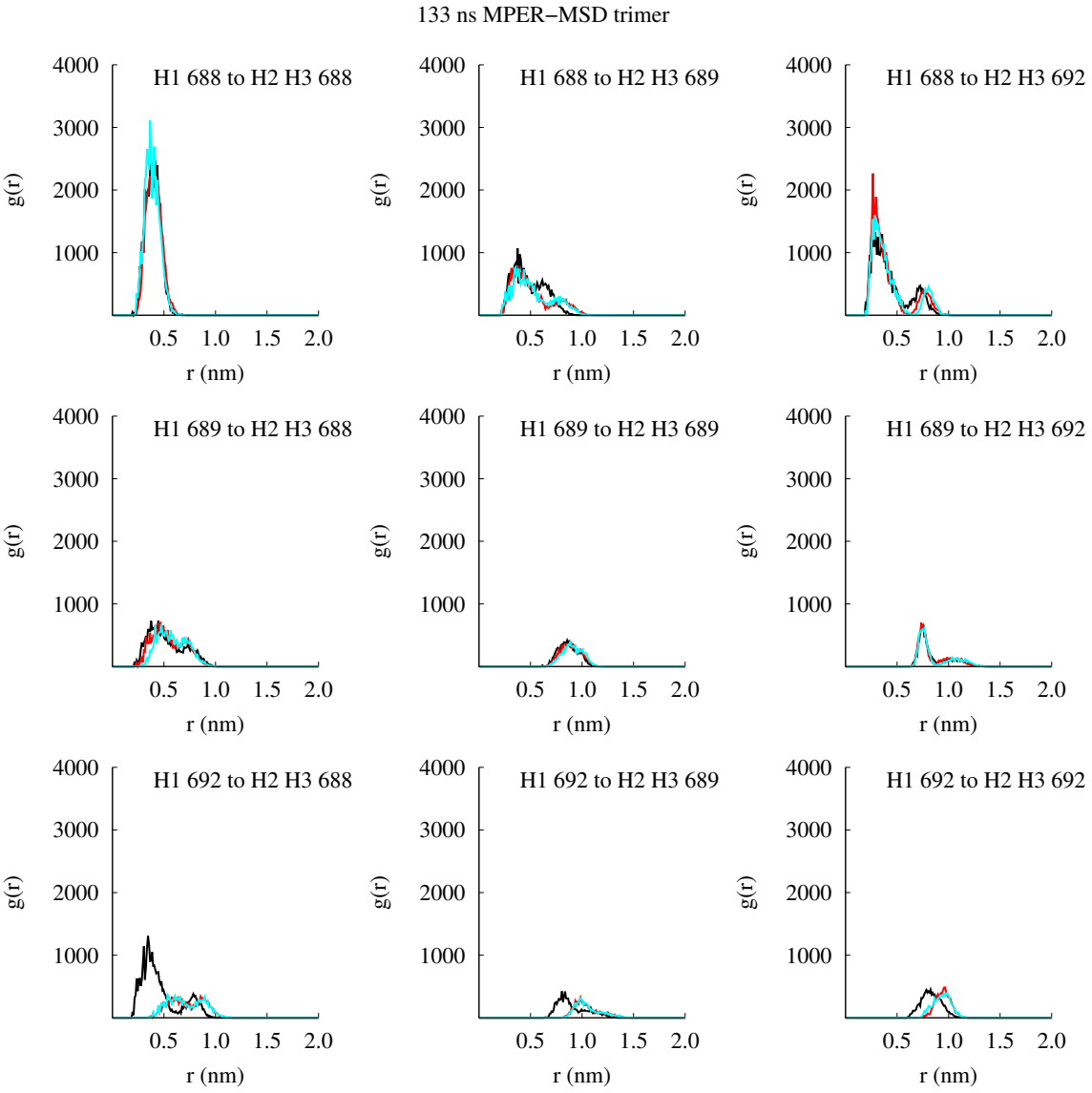

Figure 6.9: Radial distribution function vs distance in nm for 133 ns MPER-MSD trimer from Chapter 5. 0-50 ns in black, 50-100 ns in red, 100-133 ns in blue.

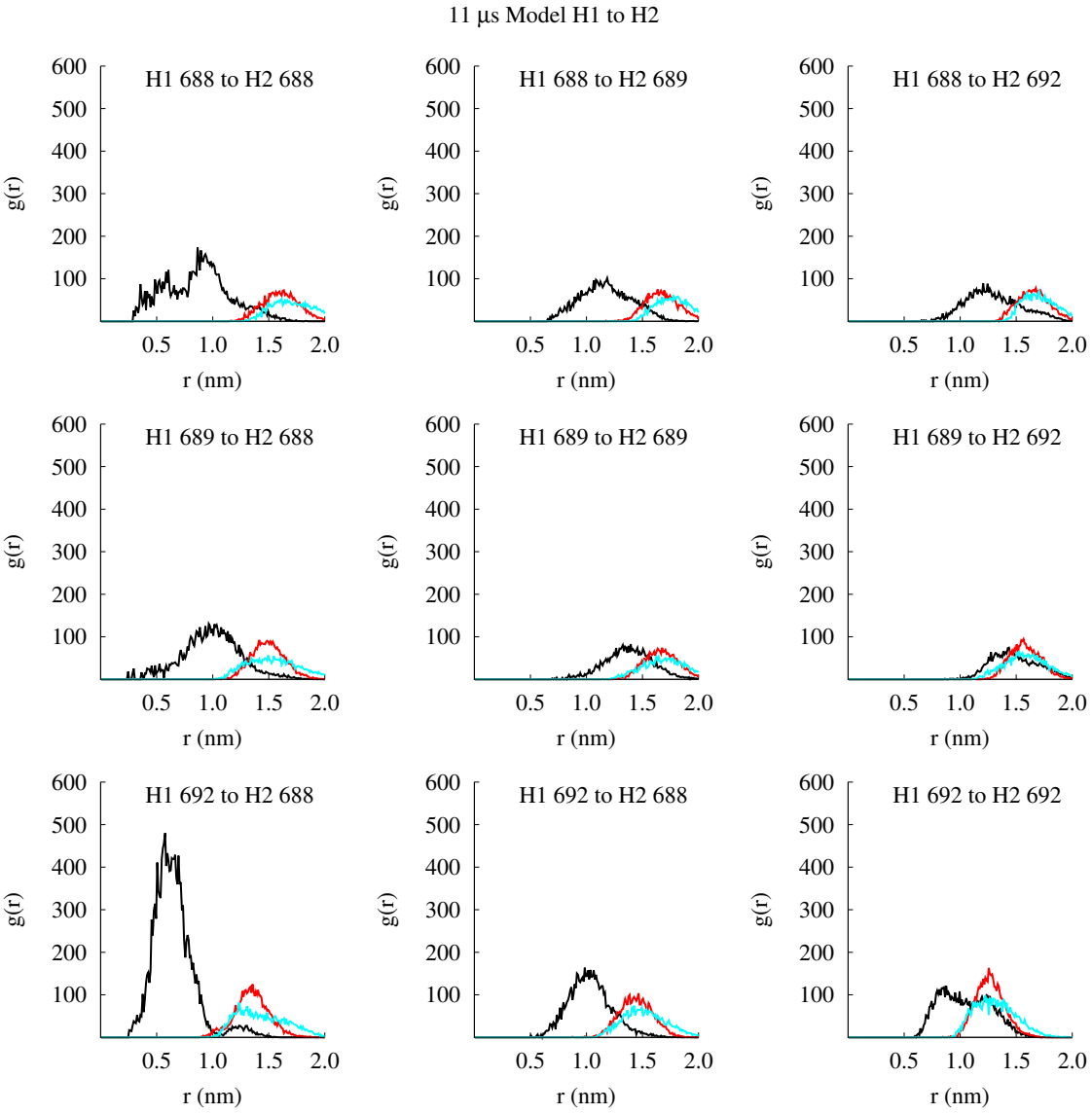

Figure 6.10: H1 to H2 radial distribution function vs distance in nm for 11 μs MPER-MSD trimer model from Chapter 5. 0-4 μs in black, 4-8 μs in red, 8-11 μs in blue.

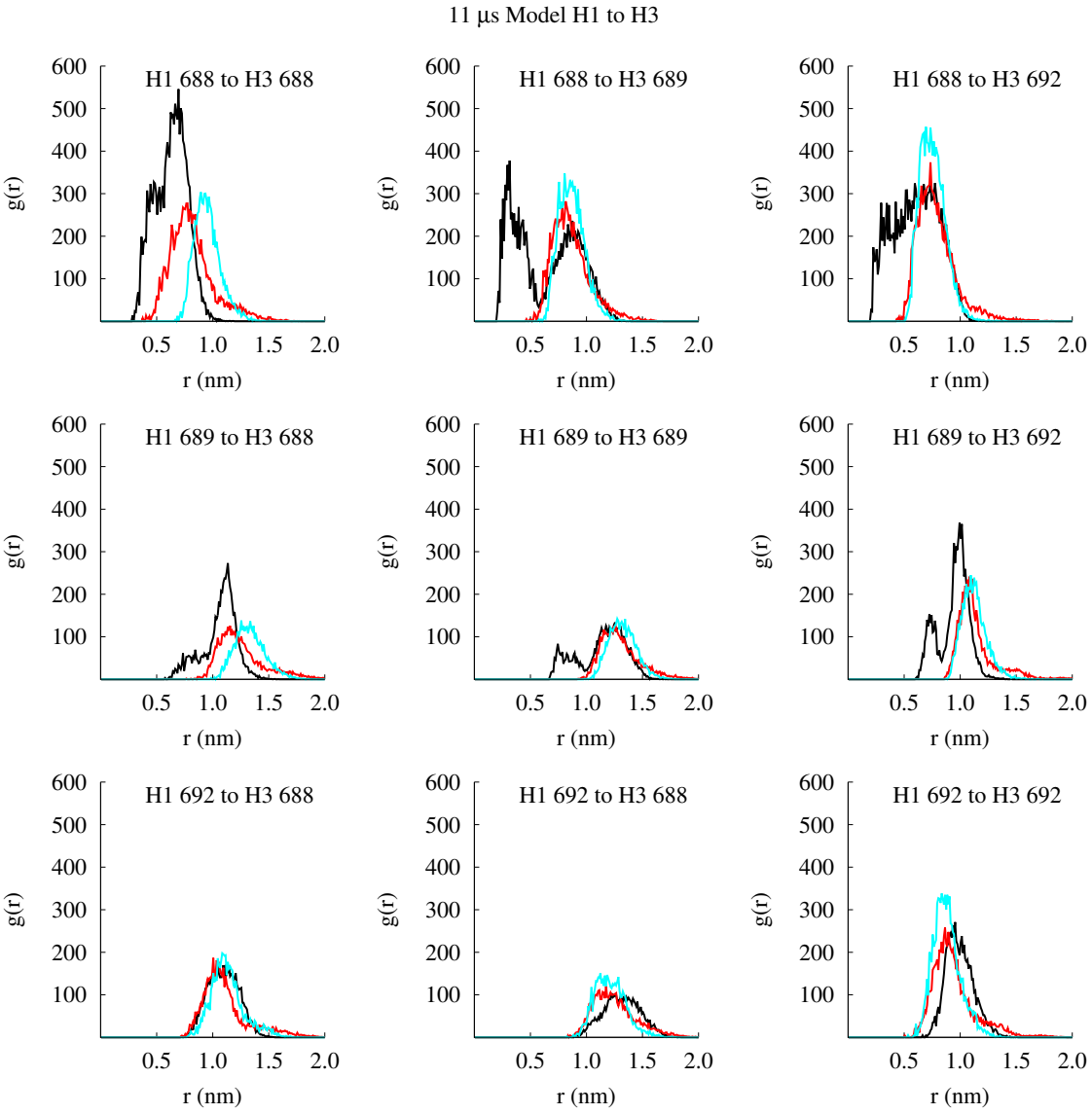

Figure 6.11: H1 to H3 radial distribution function vs distance in nm for 11 μs MPER-MSD trimer model from Chapter 5. 0-4 μs in black, 4-8 μs in red, 8-11 μs in blue.

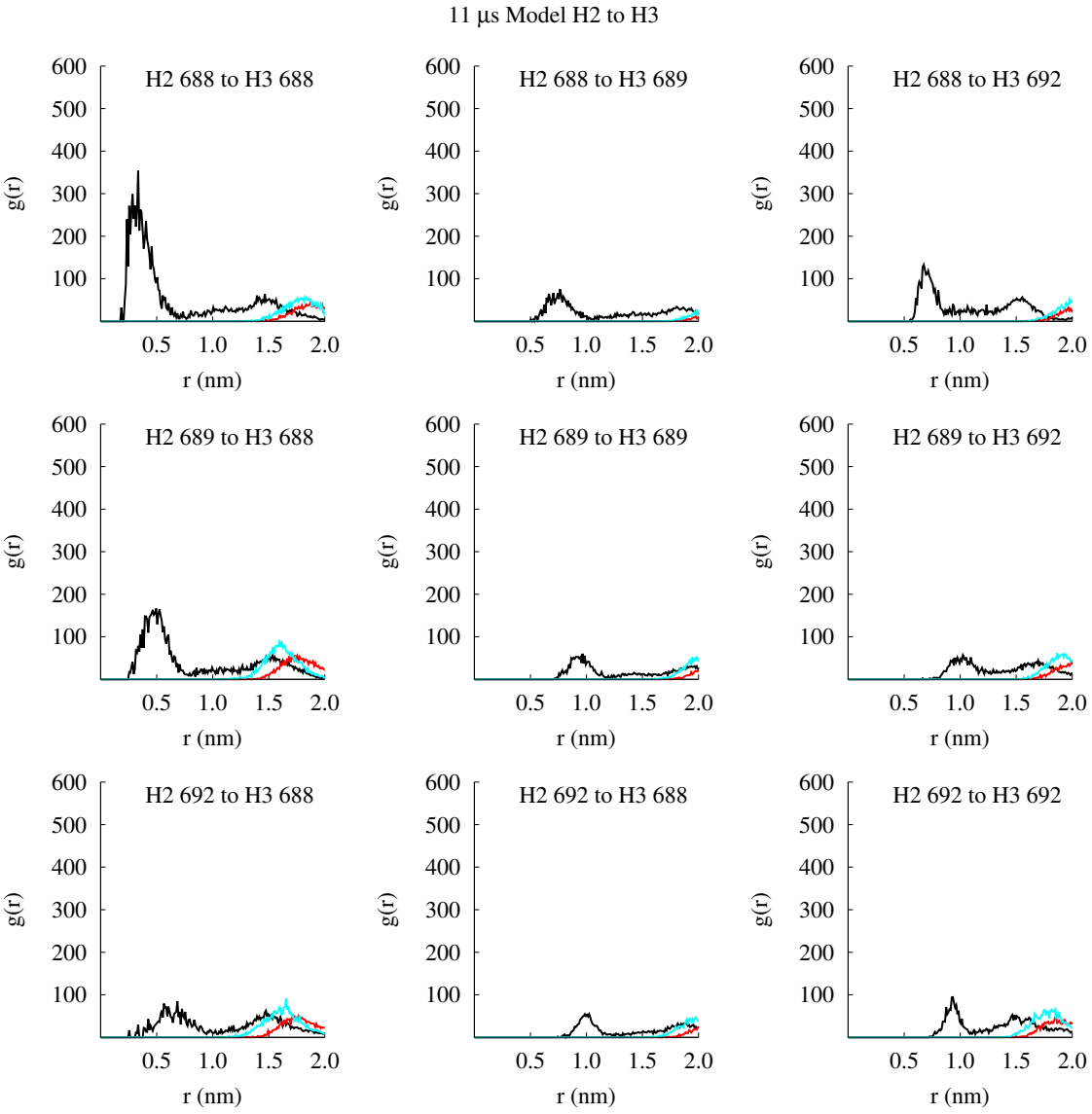

Figure 6.12: H2 to H3 radial distribution function vs distance in nm for 11 μs MPER-MSD trimer model from Chapter 5. 0-4 μs in black, 4-8 μs in red, 8-11 μs in blue.

7. Conclusion and future work

7.1 Thesis summary

In Chapter 3, Aim I characterized the solvation of the highly conserved midspan arginine of the HIV-1 gp41 MSD in cholesterol-containing and cholesterol-free bilayers using all-atom MD. The MSD was membrane spanning and stable as an α-helix during 300 ns MD in bilayers with and without cholesterol. In cholesterol-containing bilayers, the midspan arginine was solvated by a stable, but fluctuating, water defect. Thinning of the membrane and thickness fluctuations near the defect also contributed to arginine solvation. The mechanism also involved tilting the helix $\sim 20^{\circ}$ with respect to membrane normal to allow snorkeling of the arginine to lipid headgroups. Mutation of the arginine to leucine (R694L) retained the α-helical stable state (as determined by metadynamics) but exhibited no water defect and no local membrane thinning or thickness fluctuations. Mutation of R694L back to arginine showed that the midspan arginine was capable of water influx that was independent of initial condition. In cholesterol-free bilayers, the MSD was also solvated, as the water defect apparently did not depend on the presence or absence of cholesterol or the local lipid composition. However, the tilt of the MSD depended on the location of the peptide in the cholesterol-free bilayers and did not depend only on the presence of the water defect. The midspan arginine of the MSD and its water defect may represent a local membrane metastability and the consistent tilt of the helix may be important for MSD trimerization in cholesterol-containing bilayers. This work has been published in the Biochimica et Biophysica Acta - Biomembranes [144].

In Chapter 4, Aim II examined the dynamics of a model viral bilayer with 50% cholesterol. Long MD of up to 9.98 μs allowed the cholesterol-containing model

viral bilayers to achieve more configurational sampling than the 300 ns MD in Aim I. Sufficient statistics allowed calculation of the diffusion coefficients of the lipids, cholesterol molecules, and MSD peptides, which agreed with experiments. Observed temperature differences of 4 K resulted in changes in diffusion, lipid order, membrane area, and thickness. Despite simulation times of over an order of magnitude greater than previous studies, the bilayers did not achieve complete configurational sampling, due to the high concentration of cholesterol. This manuscript is in revision.

In Chapter 5, Aim III enumerated ∼270,000 configurations of trimeric MSD in vacuum and determined the lowest-energy structure to have interacting GXXXG motifs. This configuration was not stable in a bilayer for 150 ns, however the addition of the experimental crystal structure of trimeric MPER stabilized the trimer for over 100 ns. The resulting trimeric MPER-MSD in a cholesterol-containing bilayer was simulated for ∼11 μs NPT MD, allowing it to relax into a different configuration that may precede the prefusion intermediate stage during HIV-1 fusion.

In Chapter 6, Aim IV examined the stability of dimeric MSD to elucidate the possible oligomeric form of MSD. The lowest energy WT and R694L dimers from similar enumerations as in Aim III were simulated in cholesterol-containing bilayers. Unlike the trimeric MSD and MPER-MSD models, the highly conserved GXXXG motifs remained associated in the dimers. So did other initial conditions of the WT and R694L dimers. This indicates that the highly conserved glycine motif may be required during fusion for other roles besides trimerization. The manuscript encompassing Aims III and IV is in preparation.

7.2 Future computational work

The next obvious step towards building a model of the HIV-1 spike is to augment the MPER-MSD model created in this thesis with the ectodomain of gp41. A careful

choice of experimental structure with regard to attached ligands, possible stage of fusion, structure resolution, missing residues or atoms, etc. is necessary and further refinement may be needed. In Aim III, the MPER-MSD model after 11 μs represented a possible structure after prefusion but before the prefusion intermediate stage of HIV-1 infection. Free-energy simulation techniques, such as string method, between the model and the prefusion intermediate NMR of MPER can construct a low energy pathway between both structures [145]. The prefusion intermediate state is accessible to antibodies and the antibody binding process can also be modeled with MD.

As mentioned in the conclusion of Chapter 6, MD of the WT dimer, WT-LED, for $\sim$10 μs can indicate if the GGXXG residues also interact on the μs timescale, as it took several μs for the MPER-MSD trimer to relax to a different configuration where the glycine motifs did not associate between the helices. This longer simulation can add to the knowledge of the possible oligomeric form of MSD. Also, simulations of mutations of GGXXG in the dimer can more thoroughly assess the role of this motif in dimerization on 150 ns MD timescales. Simulations of a MPER-MSD dimer can assess how other segments of gp41 potentially stabilize the dimer form.

As a first step towards a more realistic model viral bilayer, future research can utilize the newly published sphingomyelin parameters [86] instead of using the saturated lipid substitute, DPPC. Also, the cellular and viral bilayer leaflets do not have an identical lipid distribution [146, 147] and this uneven distribution can be modeled with MD.

7.3 Experimental validation

Solid-state NMR spectroscopy has observed the interaction of water with an arginine from transmembrane peptides [64, 148]. Solid-state NMR has also examined the interactions of cholesterol and DPPC [87]. The simulations in Aim I can be validated

with similar experiments with WT and R694L MSD peptides in cholesterol-containing and cholesterol-free bilayers. ATR-FTIR has been used on HIV-1 FP to determine its orientation in native and cholesterol-extracted virus to determine tilt angle [149]. Perhaps this can be repeated for the MSD and compared to the tilt angles measured in Aim I.

In Aim I, the MSD was modeled in cholesterol-free bilayers to represent a noninfectious virus, since cholesterol is required for fusion. Viruses that have been treated with methyl-β-cyclodextrin to remove cholesterol are no longer infectious. However, the membrane environment local to the spikes during cholesterol extraction is unknown. If cholesterol is first removed from directly near the spikes, then the spike will exist in a more gel-like than liquid-ordered bilayer. Or, the spike could hold onto local cholesterol during methyl-β-cyclodextrin treatment and this would change the relevance of the pure DPPC simulations in Aim I. Investigations into methyl-β-cyclodextrin-mediated cholesterol removal from the virus and associated changes in membrane characteristics are currently being investigated by our collaborators at Drexel College of Medicine, Department of Biochemistry and Molecular Biology, in the lab of Dr. Irwin Chaiken.

Previous studies on the trimerization ability of HIV-1 MSD by TOXCAT supported oligomerization. Modeling of trimeric MSD asserted that trimerization of the MSD was more stable with deprotonated midspan arginines [150]. Similar studies of the severe acute respiratory syndrome (SARS) coronavirus spike protein membrane spanning domain stated that the GXXXG motif is a trimerization motif using ToxR/-TOXCAT assays and SDS-PAGE gels [151]. However, a later study that looked at the SARS CoV S TMD in the context of the full-length sequence found that SDS-PAGE of the WT and mutants of the GXXXG domain had similar ratios of oligomeric forms [152]. This indicated that the GXXXG motif did not mediate oligomerization

and that it is important to study the SARS TMD and other similar proteins in the context of the full length protein. More careful studies of the oligomerization of HIV-1 gp41 MSD with mutagenesis to the GXXXG motif need to be performed and in the context of full length gp41.

7.4 Thesis impact

The simulations in this thesis emphasized the importance of focusing first on a small segment of a complex, oligomeric, transmembrane protein and that segment's local environment (Aims I and II) before building upon that knowledge to create oligomeric structures (Aims III and IV). Specifically, this thesis asserts the following:

1. From extensive MD simulations of the excised HIV-1 gp41 MSD monomeric peptide in cholesterol-containing and cholesterol-free model viral bilayers, the mechanism of solvation of the highly conserved midspan arginine was determined to be a fluctuating water defect and the presence of cholesterol bestowed stability to the tilt of the helix with respect to membrane normal. From these simulations, we assert that cholesterol allows the spike to localize and control the membrane thinning, thickness fluctuations, and water depth fluctuations. This local metastability can then be "triggered" during fusion.

2. By configurational sampling through μs-long MD, small changes in system temperature resulted in large differences in the dynamics of a model viral bilayer. Diffusive motion was observed, despite the high percentage of cholesterol in the system, and calculated diffusion coefficients of peptides, lipids, and cholesterol molecules agreed with experiments. Model membranes with high cholesterol content (such as a model of the HIV-1 viral membrane) continue to be difficult to simulate, as the cholesterol decreases the dynamics to timescales beyond

simulations on even highly specialized, state-of-the-art supercomputers. Importantly, the water defect was stable on 10 μs and this further supports the water defect as a feature of the spike.

3. Enumerations of ~270,000 Euler transformations led to discovery of the lowest energy trimeric MSD in vacuum, which was simulated for 150 ns but required the addition of MPER to stabilize in a bilayer. The trimeric MPER-MSD model was simulated for 11 μs in a cholesterol-containing bilayer to allow relaxation into a lower energy structure. This represents the first simulation of a model of trimeric MPER-MSD in a cholesterol-containing bilayer and it is a link between the prefusion and prefusion intermediate recent experimental structures. Also, this simulation is the first to show interaction between HIV-1 MPER CRAC lysines to water and cholesterol and hopefully will inspire new insights into the unique structure/function relationships of this region of HIV-1 gp41.

4. Comparison of dimeric MSD with previous trimeric MSD simulations revealed that the GXXXG motif can more easily mediate dimerization and that perhaps it has other roles in HIV-1 fusion than trimerization. Dimerization may be needed after translation at the ER. Or, the MSD may exist as a dimer and monomer in the prefusion spike, which would support an asymmetric 6HB-formation. Also, if more than one spike is required for fusion and infection, the GXXXG motif may mediate dimerization of MSDs from one spike to another during fusion.

Bibliography

[1] Gangupomu, V. K.; Abrams, C. F. *Biophys. J.* **2010**, *99*, 3438–3444.

[2] UNAIDS, *Global Report: UNAIDS Report on the Global AIDS Epidemic 2013.*; 2013.

[3] Zhu, P.; Chertova, E.; Bess, J.; Lifson, J. D.; Arthur, L. O.; Liu, J.; Taylor, K. A.; Roux, K. H. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 15812–15817.

[4] Frankel, A. D.; Young, J. A. *Annu. Rev. Biochem.* **1998**, *67*, 1–25.

[5] Gallo, S. *Biochim. Biophys. Acta* **2003**, *1614*, 36–50.

[6] Waheed, A. A.; Freed, E. O. *Virus Res.* **2009**, *143*, 162–176.

[7] Checkley, M. A.; Luttge, B. G.; Freed, E. O. *J. Mol. Biol.* **2011**, *410*, 582–608.

[8] Brügger, B.; Glass, B.; Haberkant, P.; Leibrecht, I.; Wieland, F. T.; Kräusslich, H. G. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 2641–2646.

[9] Liao, Z.; Cimakasky, L. M.; Hampton, R.; Nguyen, D. H.; Hildreth, J. E. *AIDS Res. Hum. Retroviruses* **2001**, *17*, 1009–1019.

[10] Graham, D. R. M.; Chertova, E.; Hilburn, J. M.; Arthur, L. O.; Hildreth, J. E. K. *J. Virol.* **2003**, *77*, 8237–8248.

[11] Campbell, S. M.; Crowe, S. M.; Mak, J. *AIDS* **2002**, *16*, 2253–2261.

[12] Campbell, S.; Gaus, K.; Bittman, R.; Jessup, W.; Crowe, S.; Mak, J. *J. Virol.* **2004**, *78*, 10556–10565.

[13] Guyader, M.; Kiyokawa, E.; Abrami, L.; Turelli, P.; Trono, D. *J. Virol.* **2002**, *76*, 10356–10364.

[14] Ono, A.; Freed, E. O. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 13925–13930.

[15] Zhukovsky, M. A.; Lee, P.-H.; Ott, A.; Helms, V. *Proteins: Struct., Funct., Bioinf.* **2013**, *81*, 555–567.

[16] Sankaram, M. B.; Thompson, T. E. *Biochemistry* **1990**, *29*, 10676–10684.

[17] Contarino, M.; Bastian, A. R.; Sundaram, R. V. K.; McFadden, K.; Duffy, C.; Gangupomu, V.; Baker, M.; Abrams, C.; Chaiken, I. *Antimicrob. Agents Chemother.* **2013**, *57*, 4743–4750.

[18] Zanetti, G.; Briggs, J. A. G.; Grünewald, K.; Sattentau, Q. J.; Fuller, S. D. *PLoS Pathog.* **2006**, *2*, e83.

[19] Liu, J.; Bartesaghi, A.; Borgnia, M. J.; Sapiro, G.; Subramaniam, S. *Nature* **2008**, *455*, 109–113.

[20] Wu, S. R.; Löving, R.; Lindqvist, B.; Hebert, H.; Koeck, P.; Sjöberg, M.; Garoff, H. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 18844–18849.

[21] Chan, D. C.; Fass, D.; Berger, J. M.; Kim, P. S. *Cell* **1997**, *89*, 263–273.

[22] Kwong, P. D.; Wyatt, R.; Robinson, J.; Sweet, R. W.; Sodroski, J.; Hendrickson, W. A. *Nature* **1998**, *393*, 648–659.

[23] Schibli, D. J.; Montelaro, R. C.; Vogel, H. *Biochemistry* **2001**, *40*, 9570–9578.

[24] Sun, Z.; Oh, K. J.; Kim, M.; Yu, J.; Brusic, V.; Song, L.; Qiao, Z.; Wang, J.-H.; Wagner, G.; Reinherz, E. L. *Immunity* **2008**, *28*, 52–63.

[25] Liu, J.; Deng, Y.; Dey, A. K.; Moore, J. P.; Lu, M. *Biochemistry* **2009**, *48*, 2915–2923.

[26] Buzon, V.; Natrajan, G.; Schibli, D.; Campelo, F.; Kozlov, M. M.; Weissenhorn, W. *PLoS Pathog.* **2010**, *6*, e1000880.

[27] Pancera, M.; Majeed, S.; Ban, Y.; Chen, L.; Huang, C.-C.; Kong, L.; Kwon, Y. D.; Stuckey, J.; Zhou, T.; Robinson, J. E. et al. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 1166–1171.

[28] Mao, Y.; Wang, L.; Gu, C.; Herschhorn, A.; Désormeaux, A.; Finzi, A.; Xiang, S.; Sodroski, J. G. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 12438–12443.

[29] Wong, T. C. *Biochim. Biophys. Acta* **2003**, *1609*, 45–54.

[30] Lai, A. L.; Moorthy, A. E.; Li, Y.; Tamm, L. K. *J. Mol. Biol.* **2012**, *418*, 3–15.

[31] Lai, A. L.; Freed, J. H. *Biophys. J.* **2014**, *106*, 172–181.

[32] Postler, T. S.; Desrosiers, R. C. *J. Virol.* **2013**, *87*, 2–15.

[33] Dimitrov, A. S.; Jacobs, A.; Finnegan, C. M.; Stiegler, G.; Katinger, H.; Blumenthal, R. *Biochemistry* **2007**, *46*, 1398–1401.

[34] Song, L.; Sun, Z. J.; Coleman, K. E.; Zwick, M. B.; Gach, J. S.; Wang, J.; Reinherz, E. L.; Wagner, G.; Kim, M. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 9057–9062.

[35] Frey, G.; Peng, H.; Rits-Volloch, S.; Morelli, M.; Cheng, Y.; Chen, B. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 3739–3744.

[36] Liu, S.; Kondo, N.; Long, Y.; Xiao, D.; Iwamoto, A.; Matsuda, Z. *Retrovirol.* **2010**, *7*, 100.

[37] Sáez-Cirión, A.; Nir, S.; Lorizate, M.; Agirre, A.; Cruz, A.; Pérez-Gil, J.; Nieva, J. L. *J. Biol. Chem.* **2002**, *277*, 21776–21785.

[38] Greenwood, A. I.; Pan, J.; Mills, T. T.; Nagle, J. F.; Epand, R. M.; Tristram-Nagle, S. *Biochim. Biophys. Acta* **2008**, *1778*, 1120–1130.

[39] Apellániz, B.; Nir, S.; Nieva, J. L. *Biochemistry* **2009**, *48*, 5320–5331.

[40] Apellániz, B.; Ivankin, A.; Nir, S.; Gidalevitz, D.; Nieva, J. L. *Biophys. J.* **2011**, *101*, 2426–2435.

[41] Hartono, Y. D.; Mun, Y. Y.; Zhang, D. *Proteins: Struct., Funct., Bioinf.* **2013**, *81*, 933–944.

[42] Kyrychenko, A.; Freites, J. A.; He, J.; Tobias, D. J.; Wimley, W. C.; Ladokhin, A. S. *Biophys. J.* **2014**, *106*, 610–620.

[43] Zhang, Y.; Sagui, C. *J. Phys. Chem. B* **2014**, *118*, 69–80.

[44] Owens, R. J.; Burke, C.; Rose, J. K. *J. Virol.* **1994**, *68*, 570–574.

[45] Welman, M.; Lemay, G.; Cohen, E. A. *Virus Res.* **2007**, *124*, 103–112.

[46] Miyauchi, K.; Komano, J.; Yokomaku, Y.; Sugiura, W.; Yamamoto, N.; Matsuda, Z. *J. Virol.* **2005**, *79*, 4720–4729.

[47] Yue, L.; Shang, L.; Hunter, E. *J. Virol.* **2009**, *83*, 11588–11598.

[48] Russ, W. P.; Engelman, D. M. *J. Mol. Biol.* **2000**, *296*, 911–919.

[49] Long, Y.; Meng, F.; Kondo, N.; Iwamoto, A.; Matsuda, Z. *Protein Cell* **2011**, *2*, 369–376.

[50] Helseth, E.; Olshevsky, U.; Gabuzda, D.; Ardman, B.; Haseltine, W.; Sodroski, J. *J. Virol.* **1990**, *64*, 6314–6318.

[51] Miyauchi, K.; Curran, R.; Matthews, E.; Komano, J.; Hoshino, T.; Engelman, D. M.; Matsuda, Z. *Jpn. J. Infect. Dis.* **2006**, *59*, 77–84.

[52] Shang, L.; Yue, L.; Hunter, E. *J. Virol.* **2008**, *82*, 5417–5428.

[53] Shang, L.; Hunter, E. *J. Virol.* **2010**, *404*, 158–167.

[54] Miyauchi, K.; Curran, A. R.; Long, Y.; Kondo, N.; Iwamoto, A.; Engelman, D. M.; Matsuda, Z. *Retrovirology* **2010**, *7*, 95.

[55] West, J. T.; Johnston, P. B.; Dubay, S. R.; Hunter, E. *J. Virol.* **2001**, *75*, 9601–9612.

[56] Fleming, P. J.; Freites, J. A.; Moon, C. P.; Tobias, D. J.; Fleming, K. G. *Biochim. Biophys. Acta* **2012**, *1818*, 126–134.

[57] MacCallum, J. L.; Bennett, W. F. D.; Tieleman, D. P. *Biophys. J.* **2008**, *94*, 3393–3404.

[58] Vostrikov, V. V.; Hall, B. A.; Greathouse, D. V.; Koeppe, R. E.; Sansom, M. S. P. *J. Am. Chem. Soc.* **2010**, *132*, 5803–5811.

[59] Gumbart, J.; Roux, B. *Biophys. J.* **2012**, *102*, 795–801.

[60] Freites, J. A.; Tobias, D. J.; von Heijne, G.; White, S. H. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 15059–15064.

[61] Herrera, F. E.; Bouchet, A.; Lairion, F.; Disalvo, E. A.; Pantano, S. *J. Phys. Chem. B* **2012**, *116*, 4476–4483.

[62] Krepkiy, D.; Mihailescu, M.; Freites, J. A.; Schow, E. V.; Worcester, D. L.; Gawrisch, K.; Tobias, D. J.; White, S. H.; Swartz, K. J. *Nature* **2009**, *462*, 473–479.

[63] Li, L. B.; Vorobyov, I.; Allen, T. W. *Biochim. Biophys. Acta* **2012**, *1818*, 135–145.

[64] Li, S.; Su, Y.; Luo, W.; Hong, M. *J. Phys. Chem. B* **2010**, *114*, 4063–4069.

[65] Kowalski, M.; Potz, J.; Basiripour, L.; Dorfman, T.; Goh, W. C.; Terwilliger, E.; Dayton, A.; Rosen, C.; Haseltine, W.; Sodroski, J. *Science* **1987**, *237*, 1351–1355.

[66] Berman, P. W.; Nunes, W. M.; Haffar, O. K. *J. Virol.* **1988**, *62*, 3135–3142.

[67] Andreassen, H.; Bohr, H.; Bohr, J.; Brunak, S.; Buggae, T.; Cotterill, R. M. J.; Jacobsen, C.; Kusk, P.; Lautrup, B.; Petersen, S. B. et al. *J. Acquir. Immune Defic. Syndr.* **1990**, *3*, 615–622.

[68] Bill, R. M.; Henderson, P. J. F.; Iwata, S.; Kunji, E. R. S.; Michel, H.; Neutze, R.; Newstead, S.; Poolman, B.; Tate, C. G.; Vogel, H. *Nat. Biotechnol.* **2011**, *29*, 335–340.

[69] Chandler, D. *Introduction to Modern Statistical Mechanics*; Oxford University Press, 1987.

[70] Pathria, R. K. *Statistical Mechanics*; Butterworth-Heinemann, 1996.

[71] Frenkel, D.; Smit, B. *Understanding Molecular Simulations: From Algorithms to Applications*; Academic Press, 2001.

[72] Leach, A. *Molecular Modelling: Principles and Applications*; Prentice Hall, 2001.

[73] Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781–1802.

[74] Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33–38.

[75] Mackerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S. et al. *J. Phys. Chem. B* **1998**, *5647*, 3586–3616.

[76] Feller, S. E.; MacKerell, A. D. *J. Phys. Chem. B* **2000**, *104*, 7510–7515.

[77] Klauda, J. B.; Venable, R. M.; Freites, J. A.; O'Connor, J. W.; Tobias, D. J.; Mondragon-Ramirez, C.; Vorobyov, I.; MacKerell, A. D.; Pastor, R. W. *J. Phys. Chem. B* **2010**, *114*, 7830–7843.

[78] Pastor, R. W.; MacKerell, A. D. *J. Phys. Chem. Lett.* **2011**, *2*, 1526–1532.

[79] Lim, J. B.; Rogaski, B.; Klauda, J. B. *J. Phys. Chem. B* **2011**, *116*, 203–210.

[80] Martyna, G. J.; Tobias, D. J.; Klein, M. L. *J. Chem. Phys.* **1994**, *101*, 4177–4189.

[81] Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. *J. Chem. Phys.* **1995**, *103*, 4613.

[82] Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.

[83] Research Collaboratory for Structural Bioinformatics, RCSB Protein Data Bank. 2014; `http://www.rcsb.org`.

[84] Jo, S.; Lim, J. B.; Klauda, J. B.; Im, W. *Biophys. J.* **2009**, *97*, 50–58.

[85] Gangupomu, V. K. Molecular Simulations of HIV-1 gp41 Protein-Membrane Interactions. Ph.D. thesis, Drexel University, 2010.

[86] Venable, R. M.; Sodt, A. J.; Rogaski, B.; Rui, H.; Hatcher, E.; MacKerell, A. D.; Pastor, R. W.; Klauda, J. B. *Biophys. J.* **2014**, *107*, 134–145.

[87] Guo, W.; Kurze, V.; Huber, T.; Afdhal, N. H.; Beyer, K.; Hamilton, J. A. *Biophys. J.* **2002**, *83*, 1465–1478.

[88] McIntosh, T. J.; Simon, S. A.; Needham, D.; Huang, C. H. *Biochemistry* **1992**, *31*, 2012–2020.

[89] Chiang, Y.-W.; Costa-Filho, A. J.; Freed, J. H. *J. Phys. Chem. B* **2007**, *111*, 11260–11270.

[90] Redondo-Morata, L.; Giannotti, M. I.; Sanz, F. *Langmuir* **2012**, *28*, 12851–12860.

[91] Leekumjorn, S.; Sum, A. K. *Biochim. Biophys. Acta* **2007**, *1768*, 354–365.

[92] Schubert, T.; Schneck, E.; Tanaka, M. *J. Chem. Phys.* **2011**, *135*, 055105.

[93] Janiak, M. J.; Small, D. M.; Shipley, G. G. *Biochemistry* **1976**, *15*, 4575–4580.

[94] Wu, F.-G.; Jia, Q.; Wu, R.-G.; Yu, Z.-W. *J. Phys. Chem. B* **2011**, *115*, 8559–8568.

[95] Los Alamos National Security, LLC, HIV Databases. 2014; `http://www.hiv.lanl.gov`.

[96] Schow, E. V.; Freites, J. A.; Cheng, P.; Bernsel, A.; von Heijne, G.; White, S. H.; Tobias, D. J. *J. Membr. Biol.* **2011**, *239*, 35–48.

[97] Moon, C. P.; Fleming, K. G. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 10174–10177.

[98] Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562–12566.

[99] Laio, A.; Rodriguez-Fortea, A.; Gervasio, F. L.; Ceccarelli, M.; Parrinello, M. *J. Phys. Chem. B* **2005**, *109*, 6714–6721.

[100] Bussi, G.; Laio, A.; Parrinello, M. *Phys. Rev. Lett.* **2006**, *96*, 090601.

[101] Laio, A.; Gervasio, F. L. *Rep. Prog. Phys.* **2008**, *71*, 126601.

[102] Ryckaert, J.; Ciccotti, G.; Berendsen, H. *J. Comput. Phys.* **1977**, *23*, 327–341.

[103] Vermeer, L. S.; de Groot, B. L.; Réat, V.; Milon, A.; Czaplicki, J. *Eur. Biophys. J.* **2007**, *36*, 919–931.

[104] Tristram-Nagle, S.; Zhang, R.; Suter, R. M.; Worthington, C. R.; Sun, W. J.; Nagle, J. F. *Biophys. J.* **1993**, *64*, 1097–1109.

[105] Chiu, S. W.; Jakobsson, E.; Mashl, R. J.; Scott, H. L. *Biophys. J.* **2002**, *83*, 1842–1853.

[106] Li, L.; Vorobyov, I.; Allen, T. W. *J. Phys. Chem. B* **2013**, *117*, 11906–11920.

[107] Cournia, Z.; Ullmann, G. M.; Smith, J. C. *J. Phys. Chem. B* **2007**, *111*, 1786–1801.

[108] Jeon, J.-H.; Monne, H. M.-S.; Javanainen, M.; Metzler, R. *Phys. Rev. Lett.* **2012**, *109*.

[109] Falck, E.; Patra, M.; Karttunen, M.; Hyvönen, M. T.; Vattulainen, I. *Biophys. J.* **2004**, *87*, 1076–1091.

[110] Shaw, D. E.; Deneroff, M. M.; Dror, R. O.; Kuskin, J. S.; Larson, R. H.; Salmon, J. K.; Young, C.; Batson, B.; Bowers, K. J.; Chao, J. C. et al. *Commun. ACM* **2008**, 91–97.

[111] Klepeis, J. L.; Lindorff-Larsen, K.; Dror, R. O.; Shaw, D. E. *Curr. Opin. Struct. Biol.* **2009**, *19*, 120–127.

[112] Shan, Y.; Klepeis, J. L.; Eastwood, M. P.; Dror, R. O.; Shaw, D. E. *J. Chem. Phys.* **2005**, *122*, 54101.

[113] Lippert, R. A.; Predescu, C.; Ierardi, D. J.; Mackenzie, K. M.; Eastwood, M. P.; Dror, R. O.; Shaw, D. E. *J. Chem. Phys.* **2013**, *139*, 164106.

[114] Harvey, S. C.; Tan, R. K.-Z.; Cheatham, T. E. *J. Comput. Chem.* **1998**, *19*, 726–740.

[115] Chiu, S.-W.; Clark, M.; Subramaniam, S.; Jakobsson, E. *J. Comput. Chem.* **2000**, *21*, 121–131.

[116] Basconi, J. E.; Shirts, M. R. *J. Chem. Theory Comput.* **2013**, *9*, 2887–2899.

[117] Levine, B. G.; Stone, J. E.; Kohlmeyer, A. *J. Comput. Phys.* **2011**, *230*, 3556–3569.

[118] Scheidt, H. A.; Huster, D.; Gawrisch, K. *Biophys. J.* **2005**, *89*, 2504–2512.

[119] Filippov, A. V.; Rudakova, M. A.; Oradd, G.; Lindblom, G. *Biophyzika* **2007**, *52*, 476–485.

[120] Scherfeld, D.; Kahya, N.; Schwille, P. *Biophys. J.* **2003**, *85*, 3758–3768.

[121] Falck, E.; Róg, T.; Karttunen, M.; Vattulainen, I. *J. Am. Chem. Soc.* **2008**, *130*, 44–45.

[122] Miettinen, M. S.; Monticelli, L.; Hammaren, H.; Niemelä, P. S.; Bjelkmar, P.; Murtola, T.; Lindahl, E.; Vattulainen, I. *J. Am. Chem. Soc.* **2010**, *132*, 7574–7575.

[123] Lindahl, E.; Edholm, O. *J. Chem. Phys.* **2001**, *115*, 4938.

[124] Roark, M.; Feller, S. E. *J. Phys. Chem. B* **2009**, *113*, 13229–13234.

[125] Andoh, Y.; Okazaki, S.; Ueoka, R. *Biochim. Biophys. Acta* **2013**, *1828*, 1259–1270.

[126] Braun, A. R.; Brandt, E. G.; Edholm, O.; Nagle, J. F.; Sachs, J. N. *Biophys. J.* **2011**, *100*, 2112–2120.

[127] Apajalahti, T.; Niemelä, P.; Govindan, P. N.; Miettinen, M. S.; Salonen, E.; S-J, M.; Vattulainen, I. *Faraday Discuss.* **2010**, *144*, 411–430.

[128] Jo, S.; Rui, H.; Lim, J. B.; Klauda, J. B.; Im, W. *J. Phys. Chem. B* **2010**, *114*, 13342–13348.

[129] Choubey, A.; Kalia, R.; Malmstadt, N.; Nakano, A.; Vashishta, P. *Biophys. J.* **2013**, *104*, 2429–2436.

[130] Neale, C.; Hsu, J. C. Y.; Yip, C. M.; Pomès, R. *Biophys. J.* **2014**, *106*, L29–31.

[131] Montero, M.; Gulzar, N.; Klaric, K.-A.; Donald, J. E.; Lepik, C.; Wu, S.; Tsai, S.; Julien, J.-P.; Hessell, A. J.; Wang, S. et al. *J. Virol.* **2012**, *86*, 2930–2941.

[132] White, S. H.; Wimley, W. C. *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 319–365.

[133] Faham, S.; Yang, D.; Bare, E.; Yohannan, S.; Whitelegge, J. P.; Bowie, J. U. *J. Mol. Biol.* **2004**, *335*, 297–305.

[134] Park, Y.; Elsner, M.; Staritzbichler, R.; Helms, V. *Proteins: Struct., Funct., Bioinf.* **2004**, *585*, 577–585.

[135] Cymer, F.; Veerappan, A.; Schneider, D. *Biochim. Biophys. Acta* **2012**, *1818*, 963–973.

[136] Mao, Y.; Wang, L.; Gu, C.; Herschhorn, A.; Désormeaux, A.; Finzi, A.; Xiang, S.-H.; Sodroski, J. G. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 12438–12443.

[137] Reardon, P. N.; Sage, H.; Dennison, S. M.; Martin, J. W.; Donald, B. R.; Alam, S. M.; Haynes, B. F.; Spicer, L. D. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 1391–1396.

[138] Melnyk, R. A.; Kim, S.; Curran, A. R.; Engelman, D. M.; Bowie, J. U.; Deber, C. M. *J. Biol. Chem.* **2004**, *279*, 16591–16597.

[139] Kim, S.; Jeon, T. J.; Oberai, A.; Yang, D.; Schmidt, J. J.; Bowie, J. U. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 14278–14283.

[140] Raja, N. U.; Vincent, M. J.; Jabbar, M. A. *Virology* **1994**, *204*, 357–366.

[141] Subramaniam, S. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, E4172–E4174.

[142] Liu, J.; Deng, Y.; Li, Q.; Dey, A. K.; Moore, J. P.; Lu, M. *J. Virol.* **2010**, *84*, 201–209.

[143] Earl, P. L.; Doms, R. W.; Moss, B. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 648–652.

[144] Baker, M. K.; Gangupomu, V. K.; Abrams, C. F. *Biochim. Biophys. Acta* **2014**, *694*, 26–31.

[145] Abrams, C.; Bussi, G. *Entropy* **2013**, *16*, 163–199.

[146] Perlmutter, J. D.; Sachs, J. N. *J. Am. Chem. Soc.* **2011**, *133*, 6563–6577.

[147] Polley, A.; Vemparala, S.; Rao, M. *J. Phys. Chem. B* **2012**, *116*, 13403–13410.

[148] Hong, M.; Su, Y. *Protein Sci.* **2011**, *20*, 641–655.

[149] Shai, Y. *Biochim. Biophys. Acta* **2013**, *1828*, 2306–2313.

[150] Kim, J. H.; Hartley, T. L.; Curran, A. R.; Engelman, D. M. *Biochim. Biophys. Acta* **2009**, *1788*, 1804–1812.

[151] Arbely, E.; Granot, Z.; Kass, I.; Orly, J.; Arkin, I. T. *Biochemistry* **2006**, *45*, 11349–11356.

[152] Corver, J.; Broer, R.; van Kasteren, P.; Spaan, W. *J. Virol.* **2007**, *81*, 8352–8355.

Appendix A. Unique water molecules per residue script

This script, written in *tcl*, finds the number and standard deviation of water molecules that uniquely associate with each residue.

```
#Number of unique waters per residue
#mbaker
#added standard deviation

mol new WT1_nff.psf
mol addfile WT1_wrap_every10.dcd waitfor all
set nf [molinfo top get numframes]
set sel1 [atomselect top "water and same residue as within 4 of
   protein"]
set sel2 [atomselect top protein]
set w [atomselect top water]
set seg [lsort -unique [$w get segname]]
$w delete
set outfile [open WT1_water_unique_std.dat w]
set wat {}
for {set j 0} {$j < 27} {incr j} {
        lappend wat 0
}

for {set i 0} {$i < $nf} {incr i} {
        $sel1 frame $i
        $sel1 update
        $sel2 frame $i
        $sel2 update
        set list2 {}

        set cts [measure contacts 4 $sel1 $sel2]
        set wlist [lindex $cts 0]
        set plist [lindex $cts 1]

        foreach w $wlist p $plist {
                set sel [atomselect top "index $w" frame $i]
                set selp [atomselect top "index $p" frame $i]
                set d [measure bond "$w $p" frame $i]
                lappend list2 [list [$sel get segname] [$sel get
                   resid] $d [$selp get resid]]
                $sel delete
                $selp delete
        }
        #use lsearch to separate based on segname

        foreach s $seg {
                set list3 [lsearch -index 0 -all -inline $list2 $s]
```

```
                set list4 [lsort -index 1 -integer -increasing
                   $list3]
                set length [llength $list4]
                set listb {}
                if {$length > 0} {
                        set idx 0
                        while {$idx < $length} {
                                set low [lindex $list4 $idx 1]
                                lappend listb [lsort -index 1
                                   -unique [lsort -index 2
                                   -decreasing -real [lsearch -index
                                    1 -all -inline $list4 $low]]]
                                set idx [expr [lindex [lsearch
                                   -index 1 -all $list4 $low] end] +
                                    1]
                        }
                }
                #count unique waters for each residue
                for {set k 0} {$k < [llength $listb]} {incr k} {
                        lset wat [expr [lindex $listb $k 0 3] -681]
                           [expr [lindex $wat [expr [lindex $listb
                           $k 0 3] -681]] +1]
                }
        }
        puts "frame $i"

}

#average over all frames
set wat_avg {}
for {set m 0} {$m < 27} {incr m} {
        lappend wat_avg [expr [lindex $wat $m]/$nf.]
}
#Recalculate with standard deviation
set wat {}
set std {}
for {set j 0} {$j < 27} {incr j} {
        lappend wat 0
        lappend std 0
}

for {set i 0} {$i < $nf} {incr i} {
        $sel1 frame $i
        $sel1 update
        $sel2 frame $i
        $sel2 update
        set list2 {}
        set wat_frame {}
        for {set j 0} {$j < 27} {incr j} {
                lappend wat_frame 0
        }
```

```
set cts [measure contacts 4 $sel1 $sel2]
set wlist [lindex $cts 0]
set plist [lindex $cts 1]

foreach w $wlist p $plist {
        set sel [atomselect top "index $w" frame $i]
        set selp [atomselect top "index $p" frame $i]
        set d [measure bond "$w $p" frame $i]
        lappend list2 [list [$sel get segname] [$sel get
           resid] $d [$selp get resid]]
        $sel delete
        $selp delete
}
#use lsearch to separate based on segname

foreach s $seg {
        set list3 [lsearch -index 0 -all -inline $list2 $s]
        set list4 [lsort -index 1 -integer -increasing
           $list3]
        set length [llength $list4]
        set listb {}
        if {$length > 0} {
                set idx 0
                while {$idx < $length} {
                        set low [lindex $list4 $idx 1]
                        lappend listb [lsort -index 1
                           -unique [lsort -index 2
                           -decreasing -real [lsearch -index
                            1 -all -inline $list4 $low]]]
                        set idx [expr [lindex [lsearch
                           -index 1 -all $list4 $low] end] +
                            1]
                }
        }

        #count unique waters for each residue
        for {set k 0} {$k < [llength $listb]} {incr k} {
                lset wat [expr [lindex $listb $k 0 3] -681]
                   [expr [lindex $wat [expr [lindex $listb
                   $k 0 3] -681]] +1]
                lset wat_frame [expr [lindex $listb $k 0 3]
                   -681] [expr [lindex $wat_frame [expr [
                   lindex $listb $k 0 3] -681]] +1]
        }
}
#Add to std for this frame
for {set m 0} {$m < 27} {incr m} {
        lset std $m [expr [lindex $std $m] + ([lindex
           $wat_avg $m]-[lindex $wat_frame $m])*([lindex
           $wat_avg $m]-[lindex $wat_frame $m])]
}
puts "std frame $i"
```

```
}


#print to outfile
for {set m 0} {$m < 27} {incr m} {
        puts $outfile "[expr $m + 681] [lindex $wat_avg $m] [expr
           sqrt([lindex $std $m]/$nf.)]"
}

close $outfile
exit
```

Appendix B. Trimer enumeration script

This script in *tcl* enumerates the trimer structures from Aim III and saves them in the pdb and dcd format.

```
# 7/19/11
# mbaker

set j 0
set cosang {-45 -42.3 -39.4 -36.4 -33.1 -29.6 -25.5 -20.8 -14.6 0 14
   .6 20.8 25.5 29.6 33.1 36.4 39.4 42.3 45}

for {set i 0} {$i < 360} {incr i 10} {

        foreach t1 $cosang {

                for {set t2 -45} {$t2 < 50} {incr t2 5} {

                        for {set L 5} {$L < 26} {incr L 1} {

                        mol delete all
                        mol load psf final.psf pdb final.pdb
                        set sel1 [atomselect top "segname P1"]
                        set sel2 [atomselect top "segname P2"]
                        set sel3 [atomselect top "segname P3"]

                        ## Interior Rotation: alpha, 10 to 360 in 10
                            deg increments around z-axis
                        set cent1 [measure center $sel1]
                        $sel1 move [trans origin $cent1 axis z $i]
                        set cent2 [measure center $sel2]
                        $sel2 move [trans origin $cent2 axis z $i]
                        set cent3 [measure center $sel3]
                        $sel3 move [trans origin $cent3 axis z $i]

                        ## Tilt #1 -- beta, around x-axis
                        set cent1 [measure center $sel1]
                        set cent2 [measure center $sel2]
                        set cent3 [measure center $sel3]
                        $sel1 move [trans origin $cent1 axis {1 0 0}
                            $t1]
                        $sel2 move [trans origin $cent2 axis {1 0 0}
                            $t1]
                        $sel3 move [trans origin $cent3 axis {1 0 0}
                            $t1]

                        ## Tilt #2 --  gamma, around z-axis
                        set cent1 [measure center $sel1]
                        set cent2 [measure center $sel2]
```

```
                        set cent3 [measure center $sel3]
                        $sel1 move [trans origin $cent1 axis {0 0 1}
                            $t2]
                        $sel2 move [trans origin $cent2 axis {0 0 1}
                            $t2]
                        $sel3 move [trans origin $cent3 axis {0 0 1}
                            $t2]

                        ## Axisymmetric Rotation
                        set cent [measure center $sel1]
                        $sel1 move [trans origin $cent axis z 360]
                        set cent [measure center $sel2]
                        $sel2 move [trans origin $cent axis z 120]
                        set cent [measure center $sel3]
                        $sel3 move [trans origin $cent axis z 240]

                        ## Translation
                        $sel1 move [transoffset [list 0 -$L 0] ]
                        $sel2 move [transoffset [list [expr $L*cos(
                            .523599)] [expr $L*sin(.523599)] 0] ]
                        $sel3 move [transoffset [list [expr -$L*cos(
                            .523599)] [expr $L*sin(.523599)] 0] ]

                        ## Write pdb
                        set sel [atomselect top all]
                        set track [concat $i.$t1.$t2.$L]
                        set var [format %06i $j]
                        $sel writepdb ${var}.pdb
                        puts "$var = $track"
                        set j [expr $j + 1]
                        }
                }
        }
}

mol delete all

for {set j 0} {$j < 272916} {incr j 1} {
        set var [format %06i $j]
        animate read pdb ${var}.pdb
}

animate write dcd set_trimers.dcd waitfor all top

exit
```

Appendix C. Anton configuration parameters

Below are the configuration parameters for WT1, WT2, and WT3 from Aim II used with Anton software version 2.6.4.

```
anton = {
    chem = {
        Escale = "1000."
        FmaxSeparation = "2.0"
        Fscale = "3000."
        Tempmax = "350"
        maxBondLength = "4.0"
        max_strain = "0.1"
        r_over_sigmamin = "0.60"
        rmin = "0.9"
    }
    run_options = {
        cm_port = "37907"
        ganglia_addr = "log.anton.desres.deshaw.com"
        jobid = "105239"
        jobstep_id = "105239"
    }
    tune = {
        BumpyBoxDelta = "1.535"
        BumpyN = "8"
        DistNumSubboxesX = "1"
        DistNumSubboxesY = "1"
        DistNumSubboxesZ = "2"
        NumSubboxesX = "1"
        NumSubboxesY = "2"
        NumSubboxesZ = "2"
        PacketsPerSubboxPerCore = "5"
        check_overflow = "true"
        checkpoint = {
            first = "0"
            interval = "1200"
            on_last_timestep = "true"
            outdir = "checkpoint.atr"
        }
        go_verbosity = "0"
        jobstep_wallclock = "30"
        last_time = "Inf"
        machine_size = ["8" "8" "8"]
        optional = {
            AllowDRAMMode = "false"
            BondChunksPerNode = "1"
        }
        regenerateBondProgram = "true"
        trajectory = {
```

```
            first = "0"
            format = "dtr"
            interval = "240"
            outdir = "run.atr"
            periodicfix = "true"
            write_velocity = "true"
        }
    }
}
boot = {
    file = "WT_final.dms"
}
force = {
    nonbonded = {
        electrostatic = {
            type = "ewald"
        }
        far = {
            n_k = ["64" "64" "64"]
            r_spread = "6.654"
            sigma = "1.9987426757812492"
            sigma_s = "1.3295978613281245"
            type = "gse"
        }
        r_cut = "9.0"
        r_lazy = "9.5"
    }
}
global_cell = {
    r_clone = "7.008"
}
integrator = {
    Multigrator = {
        barostat = {
            MTK = {
                T_ref = "310"
                tau = "0.0416667"
                thermostat = {
                    NoseHoover = {
                        chain = {
                            mts = "1"
                            tau = ["0.0416667" "0.0416667" "0
                                .0416667"]
                        }
                    }
                    type = "NoseHoover"
                }
            }
            interval = "240"
            type = "MTK"
        }
        nve = {
```

```
            PLS = {
                number_of_stages = "3"
            }
            type = "Verlet"
        }
        thermostat = {
            NoseHoover = {
                chains = [{
                    mts = "1"
                    tau = ["0.0416667" "0.0416667" "0.0416667"]
                }]
                use_molecular_ke = "true"
            }
            interval = "24"
            type = "NoseHoover"
        }
    }
    dt = "0.0020"
    pressure = {
        isotropy = "semi_isotropic"
        max_margin_contraction = "0.9"
        p_ref = "1"
    }
    remove_com_motion = "false"
    respa = {
        bonded_interval = "1"
        nonbonded_far_interval = "3"
        nonbonded_near_interval = "1"
    }
    temperature = [{
        T_ref = "310"
    }]
    type = "Multigrator"
}
```

Below are the configuration parameters for MPER-MSD from Aim III used with Anton software version 2.12.4.

```
anton = {
    chem = {
        Escale = "3000.0000"
        FmaxSeparation = "2.0"
        Fscale = "4000.0000"
        Tempmax = "350"
        average_dispersion_type = "manual"
        maxBondLength = "4.0"
        max_strain = "0.1"
        r_over_sigmamin = "0.60"
        rmin = "0.9"
    }
    run_options = {
        cm_port = "57602"
```

```
        ganglia_addr = "log.anton.desres.deshaw.com"
        jobid = "263210"
        jobstep_id = "263210"
    }
    tune = {
        BumpyBoxDelta = "1.293"
        BumpyN = "4"
        DistNumSubboxesX = "2"
        DistNumSubboxesY = "1"
        DistNumSubboxesZ = "2"
        NumSubboxesX = "2"
        NumSubboxesY = "1"
        NumSubboxesZ = "2"
        PacketsPerSubboxPerCore = "7"
        check_overflow = "true"
        checkpoint = {
            first = "0"
            interval = "12000"
            on_last_timestep = "true"
            outdir = "checkpoint.atr"
        }
        go_verbosity = "0"
        jobstep_wallclock = "30"
        last_time = "Inf"
        machine_size = ["8" "8" "8"]
        optional = {
            BondChunksPerNode = "1"
        }
        regenerateBondProgram = "true"
        trajectory = {
            first = "0"
            format = "dtr"
            interval = "240"
            outdir = "run.dtr"
            periodicfix = "true"
        }
    }
}
boot = {
    file = "trimer_para.dms"
}
force = {
    nonbonded = {
        average_dispersion = "0"
        electrostatic = {
            type = "ewald"
        }
        far = {
            n_k = ["64" "64" "64"]
            r_spread = "7.457"
            sigma = "2.3823378906249997"
            sigma_s = "1.5761651171874997"
```

```
            type = "gse"
        }
        r_cut = "10.78"
        r_lazy = "11.28"
    }
}
global_cell = {
    r_clone = "7.811"
}
integrator = {
    Multigrator = {
        barostat = {
            MTK = {
                T_ref = "310.00"
                tau = "0.0416667"
                thermostat = {
                    NoseHoover = {
                        chain = {
                            mts = "1"
                            tau = ["0.0416667" "0.0416667" "0
                                .0416667"]
                        }
                    }
                    type = "NoseHoover"
                }
            }
            interval = "240"
            type = "MTK"
        }
        nve = {
            PLS = {
                number_of_stages = "3"
            }
            type = "Verlet"
        }
        thermostat = {
            Langevin = {
                seed = "0x98765432"
                tau = "0.3"
                use_molecular_ke = "false"
            }
            interval = "24"
            type = "Langevin"
        }
    }
    dt = "0.0020"
    pressure = {
        isotropy = "semi_isotropic"
        max_margin_contraction = "0.9"
        p_ref = "1"
        tension_ref = ["0" "0" "0" "0" "0" "0" "0" "0" "0"]
    }
```

```
    remove_com_motion = "true"
    respa = {
        bonded_interval = "1"
        nonbonded_far_interval = "3"
        nonbonded_near_interval = "1"
    }
    temperature = [{
        T_ref = "310"
    }]
    type = "Multigrator"
}
```

Vita

Michelle Katherine Baker was born on May 5, 1984 in Atlantic City, New Jersey, USA. Michelle attended Georgia Institute of Technology and graduated with a Bachelor of Science in Chemical and Biomolecular Engineering in August 2007. She also received a Master of Science in Chemical Engineering from Rowan University in August 2009, after completing thesis research on extraction of fluorescent proteins using reverse micelles. Michelle started her PhD at Drexel University in August 2009 in the Department of Chemical and Biological Engineering and joined the lab of Prof. Cameron Abrams in December. Michelle has one first-author publication [144], one in revision, and one in preparation. She plans to pursue an R&D position in industry.